Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US04/022959

International filing date: 16 July 2004 (16.07.2004)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/487,906
Filing date: 17 July 2003 (17.07.2003)

Date of receipt at the International Bureau: 10 September 2004 (10.09.2004)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

1220042

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

September 01, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/487,906*
FILING DATE: *July 17, 2003*
RELATED PCT APPLICATION NUMBER: *PCT/US04/22959*

Certified by

Jon W Dudas

Acting Under Secretary of Commerce
for Intellectual Property
and Acting Director of the U.S.
Patent and Trademark Office

PTO/SB/16 (10-01)
Approved for use through10/31/2002. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No.	EV 326482161 US

INVENTOR(S)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Parry John	Guilford	Dunedin, New Zealand

☐ *Additional inventors are being named on the _____ separately numbered sheets attached hereto*

TITLE OF THE INVENTION (500 characters max)

Markers For Detection of Gastric Cancer

Direct all correspondence to: **CORRESPONDENCE ADDRESS**

☑ Customer Number 23910 → *Place Customer Number Bar Code Label here*

OR *Type Customer Number here*

☐ Firm or Individual Name	D. Benjamin Borson, Ph.D., Esq.				
Address	Fliesler Dubb Meyer & Lovejoy LLP				
Address	4 Embarcadero Center, Ste. 400				
City	SF	State	CA	ZIP	94111
Country	USA	Telephone	415.362.3800	Fax	415.382.2928

ENCLOSED APPLICATION PARTS *(check all that apply)*

☑ Specification *Number of Pages* 21
☑ Drawing(s) *Number of Sheets* 31
☐ Application Data Sheet. See 37 CFR 1.76
☐ CD(s), Number
☐ Other (specify)

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT

☑ Applicant claims small entity status. See 37 CFR 1.27.
☑ A check or money order is enclosed to cover the filing fees
☑ The Commissioner is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: 06-1325
☐ Payment by credit card. Form PTO-2038 is attached.

FILING FEE AMOUNT ($): $80.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☑ No.
☐ Yes, the name of the U.S. Government agency and the Government contract number are: _____

Respectfully submitted,

SIGNATURE D. Benjamin Borson

Date 07/17/2003

TYPED or PRINTED NAME D. Benjamin Borson, Ph.D.

REGISTRATION NO. *(if appropriate)* 42,349

Docket Number: PEBL-1006US0

TELEPHONE 415.362.3800

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

This collection of information is required by 37 CFR 1.51. The information is used by the public to file (and by the PTO to process) a provisional application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 8 hours to complete, including gathering, preparing, and submitting the complete provisional application to the PTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, Washington, D.C. 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Box Provisional Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

UNITED STATES PROVISIONAL PATENT APPLICATION FOR

MARKERS FOR DETECTION OF GASTRIC CANCER

Inventor: Parry Guilford

CERTIFICATE OF MAILING BY "EXPRESS MAIL"
UNDER 37 C.F.R. § 1.10

"Express Mail" mailing label number: EV 326 482 161 US
Date of Mailing: July 17, 2003

I hereby certify that this correspondence is being deposited with the United States Postal Service, utilizing the "Express Mail Post Office to Addressee" service addressed to **BOX PROVISIONAL PATENT APPLICATION, Assistant Commissioner for Patents, Box 1450, Arlington Virginia** and mailed on the above Date of Mailing with the above "Express Mail" mailing label number.


(Signature)

D. Benjamin Borson, Reg. No. 42,349
Signature Date: July 17, 2003

MARKERS FOR DETECTION OF GASTRIC CANCER

Field of the Invention

This invention relates to detection of cancer. Specifically, this invention relates to the use of genetic markers for detection of cancer, and more particularly to the use of genetic markers for detection of gastric cancer.

BACKGROUND

Introduction

Survival of cancer patients is greatly enhanced when the cancer is detected and treated early. In the case of gastric cancer, patients diagnosed with early stage disease have 5-year survival rates of 90%, compared to approximately 10% for patients diagnosed with advanced disease. However, the vast majority of gastric cancer patients currently present with advanced disease. Therefore, developments that lead to early diagnosis of gastric cancer can lead to an improved prognosis for the patients.

Identification of specific cancer-associated markers in biological samples, including body fluids, for example, blood, urine, peritoneal washes and stool extracts can provide a valuable approach for the early diagnosis of cancer, leading to early treatment and improved prognosis. Specific cancer markers also can provide a means for monitoring disease progression, enabling the efficacy of surgical, radiotherapeutic and chemotherapeutic treatments to be tracked. However, for a number of major cancers, the available markers suffer from insufficient sensitivity and specificity. For example, the most frequently used markers for gastric cancer, ca19-9, ca72-4 and CEA detect only about 15-50% of gastric tumors of any stage, declining to approximately 2-11% for early stage disease. Thus, there is a very high frequency of false negative tests that can lead patients and health care practitioners to believe that no disease exists, whereas in fact, the patient may have severe cancer that needs immediate attention. Moreover, these markers can give false positive signals in up to 1/3 of individuals affected by benign gastric disease.

SUMMARY OF THE INVENTION

Thus, there is an acute need for better methods for detecting the presence of cancer. This invention provides methods, compositions and devices that can provide more sensitive detection of early stage cancer, while simultaneously decreasing the frequency of false positives and false negative test results.

In certain embodiments, molecular analysis can be used to identify genes that are highly and selectively expressed in gastric tumor tissue compared to non-malignant gastric tissue. Such analyses include microarray and quantitative polymerase chain reaction (qPCR) methods. Cancer genes and proteins encoded by those genes are herein termed gastric tumor markers (GTM). It is to be understood that the term GTM does not require that the marker be specific only for gastric tumors. Rather, expression of GTM can be increased in other types of tumors, including malignant tumors. Rather a cancer gene may be expressed in a variety of cancers, including gastric, bladder, colorectal, pancreatic, ovarian, skin (e.g., melanomas), liver, esophageal, endometrial and brain cancers, among others.

In certain embodiments, microarray methods can be used to detect patterns of over-expression of one or more genes associated with cancer.

In other embodiments, quantitative polymerase chain reaction (qPCR) can be used to identify the presence of markers in tumor or other biological samples.

Selected genes that encode proteins can be secreted by or cleaved from the cell. These proteins, either alone or in combination with each other, have utility as serum or body fluid markers for the diagnosis of gastric cancer or as markers for monitoring the progression of established disease. Detection of protein markers can be carried out using methods known in the art, and include the use of monoclonal antibodies, polyclonal antiesera and the like.

BRIEF DESCRIPTION OF THE FIGURES

This invention is described with reference to specific embodiments thereof and with reference to the figures, in which:

Figure 1 depicts a table of markers and oligonucleotide sequences of markers for gastric cancer of this invention.

Figure 2 depicts a table of results obtained of studies carried out using microarray methods.

Figure 3 depicts a table of results obtained of studies carried out using quantitative PCR.

Figures 4a – 4d depict relationships between log2 fold results obtained using array and qPCR methods, in which the data is centered on the median normal for four gastric cancer markers. Grey squares correspond to non-malignant ("normal") samples and black triangles to tumor samples. Figure 4a: ASPN. Figure 4b: SPP1. Figure 4c: SPARC. Figure 4d: MMP12.

Figures 5a-5w depict histograms showing the relative frequency vs. log2 fold change data obtained from oligonucleotide microarray studies of various tumor markers. Figure 5a: ASPN; 5b: CST1,2 & 4; 5c: CSPG2; 5d: IGFBP7; 5e: INHBA; 5f: LOXL2; 5g: Lumican; 5h: SFRP4; 5i: SPARC; 5j: SPP1; 5k: THBS2; 5l: TIMP1; 5m: adlican; 5n: PRS11; 5o: ASAH1; 5p: SFRP2; 5q: GGH; 5r: MMP12; 5s: KLK10; 5t: LEPRE1; 5u:TG; 5v: EFEMP2 and 5w: TGFBI.

Figure 6 is a histogram showing the number of markers with a higher expression than the 95th percentile of the median normal expression. Results are based on qPCR data and are shown separately for each tumor sample.

DETAILED DESCRIPTION

Markers for detection and evaluation of tumors including gastric are provided. We have surprisingly found that numerous genes and proteins are associated with gastric tumors. Detection of gene products (e.g., oligonucleotides such as mRNA) and proteins and peptides translated from such oligonucleotides therefore can be used to diagnose tumors, such as gastric tumors. Array analysis of samples taken from patients with gastric tumors and from subjects without gastric tumors has led us to the surprising discovery that in many gastric tumors, specific patterns of gene expression are associated with the disease.

Cancer markers can also be detected using antibodies raised against cancer markers.

By analyzing the presence and amounts of expression of a plurality of cancer markers can thus increase the sensitivity of diagnosis while decreasing the frequency of false positive and/or false negative results.

General Approaches to Cancer Detection

The following approaches are non-limiting methods that can be used to detect cancer including gastric cancer using GTM family members.

- Microarray approaches using oligonucleotide probes selective for products of GTM genes.
- Real-time PCR on tumor samples and normal samples using marker specific primers and probes.
- Enzyme-linked immunological assays (ELISA).
- Immunohistochemistry using anti-marker antibodies on gastric tumors and lymph node metastases.
- Immunohistochemistry using anti-marker antibodies on other tumors including but not limited to colorectal, pancreatic, ovarian, melanoma, liver, esophageal, bladder, endometrial, and brain.
- Immunodetection of marker family members in sera from gastric cancer patients taken before and after surgery to remove the tumor.
- Immunodetection of marker family members in sera from healthy individuals and individuals with non-malignant diseases such as gastritis, gastric metaplasia and dysplasia.
- Immunodetection of marker family members in patients with other cancers including but not limited to colorectal, pancreatic, ovarian, melanoma, liver, oesophageal, bladder, endometrial, and brain.
- Immunodetection of marker family members in gastric fluid, peritoneal washes and stool from gastric cancer patients.
- Analysis of array or qPCR data using computers. Primary data is collected and fold change analysis is performed by comparison of levels of gastric tumor gene expression with expression of the same genes in non-tumor tissue. A threshold for concluding that expression is increased is provided (e.g., 1.5 x increase, 2-fold

increase, and in alternative embodiments, 3-fold increase, 4-fold increase or 5-fold increase). It can be appreciated that other thresholds for concluding that increased expression has occurred can be selected without departing from the scope of this invention. Further analysis of tumor gene expression includes matching those genes exhibiting increased expression with expression profiles of known gastric tumors to provide diagnosis of tumors.

In certain aspects, this invention provides methods for detecting cancer, comprising:

(a) providing a biological sample; and

(b) detecting the over expression of a GTM family member in said sample.

In other aspects, the invention includes a step of detecting over expression of GTM mRNA.

In other aspects, the invention includes a step of detecting over expression of a GTM protein.

In yet further aspects, the invention includes a step of detecting over expression of a GTM peptide.

Other aspects of this invention include a polyclonal antiserum and monoclonal antibodies specific for a GTM.

In still further aspects, the invention includes a device for detecting a GTM, comprising:

a substrate having a GTM capture reagent thereon; and

a detector associated with said substrate, said detector capable of detecting a GTM associated with said capture reagent, wherein the capture reagent includes an oligonucleotide or an antibody.

Additional aspects include kits for detecting cancer, comprising:

a substrate;

a GTM capture reagent, including one or more of a GTM-specific oligonucleotide and a GTM-specific antibody; and

instructions for use.

Yet further aspects of the invention include an expression vector for a GTM, comprising:

a promoter;

one or more enhancers;

an initiation codon;

an oligonucleotide sequence of a GTM in an open reading frame;

a stop codon; and

optionally comprising a selectable marker.

Additional aspects of this invention comprise acell for expressing a GTM comprising:

a cell capable of sustained growth in *in vitro* conditions; and

a functional expression vector containing a GTM oligonucleotide sequence, said cell being a prokaryotic cell or a eukaryotic cell.

Additional aspects of this invention include a method for manufacturing a monoclonal antibody, comprising the steps of:

immunizing a host organism with a GTM protein or fragment;

isolating a spleen cell from said host;

fusing said spleen cell with a cell capable of being propagated *in vitro* thereby producing a fused cell;

selecting and isolating said fused cell; and

producing an culture from said isolated fused cells.

In yet further aspects, this invention includes a method for detecting gastric cancer, comprising the steps of:

providing a sample from a patient suspected of having gastric cancer;

measuring the presence of a GTM protein using an ELISA method.

As described herein, detection of tumors can be accomplished by measuring expression of one or more tumor-specific markers. We have unexpectedly found that the association between increased expression of GTMs and the presence of diagnosed gastric cancer is extremely high. The least significant association detected had a p value of about 1.6×10^{-6}. Many of the associations were significant at p values of less than 10^{-20}. With such a high significance, it may not be necessary to detect increased expression in

more than one GTM. However, the redundancy in the GTMs of this invention can permit detection of gastric cancers with an increased reliability.

The methods provided herein also include assays of high sensitivity. qPCR is extremely sensitive, and can be used to detect gene products in very low copy number (e.g., 1 – 100) in a sample. With such sensitivity, very early detection of events that are associated with gastric cancer is made possible.

EXAMPLES

The examples described herein are for purposes of illustrating embodiments of the invention. Other embodiments, methods and types of analyses are within the scope of persons of ordinary skill in the molecular diagnostic arts and need not be described in detail hereon. Other embodiments within the scope of the art are considered to be part of this invention.

Example 1: Tumor Collection

Gastric tumor samples and non-malignant gastric tissues were collected from surgical specimens resected at Seoul National University Hospital, Korea. Diagnosis of gastric cancer was made on the basis of patient history, symptoms, physical findings and histological examination of tissues.

Example 2: RNA Extraction

In some embodiments, expression of genes associated with gastric tumors were analyzed by determining the changes in RNA from samples taken from tumors. Frozen surgical specimens were embedded in OCT medium. 60μm sections were sliced from the tissue blocks using a microtome, homogenized in a TriReagent: water (3:1) mix, then chloroform extracted. Total RNA was then purified from the aqueous phase using the RNeasy™ procedure (Qiagen). RNA was also extracted from 16 cancer cell lines and pooled to serve as a reference RNA.

Example 3: Microarray Slide Preparation

Epoxy coated glass slides were obtained from MWG Biotech AG, Ebersberg, Germany) and were printed with ~30,000 50mer oligonucleotides using a Gene Machines microarraying robot, according to the manufacturer's protocol. Reference numbers (MWG oligo #) for relevant oligonucleotides, and the NCBI mRNA and protein reference sequences are shown in Figure 2.

Example 4: RNA labeling and hybridization

cDNA was transcribed from 10μg total RNA using Superscript II reverse transcriptase (Invitrogen) in reactions containing 5-(3-aminoallyl)- 2' deoxyuridine –5'-triphosphate. The reaction was then de-ionised in a Microcon column before being incubated with Cy3 or Cy5 in bicarbonate buffer for 1 hour at room temperature. Unincorporated dyes were removed using a Qiaquick column (Qiagen) and the sample concentrated to 15ul in a SpeedVac. Cy3 and Cy5 labeled cDNAs were then mixed with Ambion ULTRAhyb buffer, denatured at 100°C for 2 mins and hybridized to the microarray slides in hybridisation chambers at 42°C for 16 hours. The slides were then washed and scanned twice in an Axon 4000A scanner at two power settings to yield primary fluorescence data on gene expression.

Example 5: Normalization Procedure

To compare expression of cancer genes from tumors and non-cancerous tissues, median fluorescence intensities detected by Genepix™ software were corrected by subtraction of the local background fluorescence intensities. Spots with a background corrected intensity of less than zero were excluded. To facilitate normalization, intensity ratios and overall spot intensities were log-transformed. Log-transformed intensity ratios were corrected for dye and spatial bias using local regression implemented in the LOCFIT™ package. Log-transformed intensity ratios were regressed simultaneously with respect to overall spot intensity and location. The residuals of the local regression provided the corrected log-fold changes. For quality control, ratios of each normalized microarray were plotted with respect to spot intensity and localization. The plots were subsequently visually inspected for possible remaining artifacts. Additionally, an analysis

of variance (ANOVA) model was applied for the detection of pin-tip bias. All results and parameters of the normalization were inserted into a Postgres-database for statistical analysis.

Example 6: Statistical Analysis

Statistically significant changes in gene expression in tumor samples vs. normal tissues were identified by measured fold changes between arrays. To accomplish this, log2 (ratios) were scaled to have the same overall standard deviation per array. This standardization procudure reduced the average within-tissue class variability. The log2 (ratios) were further shifted to have a median value of zero for each oligonucleotide to facilitate visual inspection of results. A rank-test based on fold changes was then used to improve the noise robustness. This test consisted of two steps: (i) calculation of the rank of fold change (Rfc) within arrays and ii) subtraction of the median (Rfc) for normal tissue from the median(Rfc) for tumor tissue. The difference of both median ranks defines the score of the fold change rank presented in **Figure 2**.Two additional statistical tests were also performed on this standardized data: 1) Two sample student's t-test, with and without the Bonferroni adjustment and 2) the Wilcoxon test.

Example 7: Quantitative Real-Time PCR

In other embodiments, real-time, or quantitative PCR (qPCR) can be used for absolute or relative quantitation of PCR template copy number. Taqman™ probe and primer sets were designed using Primer Express V 2.0™ (Applied Biosystems). Where possible, all potential splice variants were included in the resulting amplicon, with amplicon preference given to regions covered by the MWG-Biotech-derived microarray oligonucleotide. Alternatively, if the target gene was represented by an Assay-on-Demand™ expression assay (Applied Biosystems) covering the desired amplicons, these were used. The name of the gene, symbol, the Applied Biosystems "assay on demand" number, forward primer, reverse primer and probe sequence used for qPCR are shown. In the in-house designed assays, primer concentration was titrated using a SYBR green labeling protocol and cDNA made from the reference RNA. Amplification was carried out on an ABI Prism™ 7000 sequence detection system under standard cycling

conditions. When single amplification products were observed in the dissociation curves, standard curves were generated over a 625-fold concentration range using optimal primer concentrations and 5'FAM - 3'TAMRA phosphate Taqman™ probe (Proligo) at a final concentration of 250nM. Assays giving standard curves with regression coefficients over 0.98 were used in subsequent assays. It can be appreciated that in other embodiments, regression coefficients need not be as high. Rather, any standard curve can be used so long as the regression coefficients are sufficiently high to permit statistically significant determination of differences in expression. Such regression coefficients may be above about 0.7, above about 0.8, above about 0.9 or above about 0.95 in alternative embodiments.

Assays were performed over two 96 well plates with each RNA sample represented by a single cDNA. Each plate contained a reference cDNA standard curve, over a 625-fold concentration range, in duplicate. Analysis consisted of calculating the ΔCT (target gene CT – mean reference cDNA CT). ΔCT is directly proportional to the negative log2 fold change. Log2 fold changes relative to the median non-malignant log2 fold change were then calculated (log2 fold change – median normal log2 fold change). These fold changes were then clustered into frequency classes and graphed.

Example 8: Microarray Analysis of Cancer Marker Genes

RNA from 58 gastric tumors and 58 non-malignant ("normal") gastric tissue samples were labeled with Cy5 and hybridized in duplicate or triplicate with Cy3 labeled reference RNA. After normalization, the change in expression in each of 29,718 genes was then estimated by three measures: (i) fold change: the ratio of the gene's median expression (unstandardized) in the tumor samples divided by the median level in the non-malignant samples. (ii) fold change rank and (iii) the statistical probability that the observed fold changes were significant.

Selection of Serum Markers for Gastric Malignancy

In certain embodiments, the cancer marker can be found in biological fluids, including serum. Serum markers were selected from the array data based on (i) likelihood that the encoded protein is secreted from the cell or cleaved from the

membrane. (ii) the median level of over-expression (fold change) in tumors compared to non-malignant controls, (iii) the median change in expression rank between tumors and non-malignant controls, and (iv) the degree of overlap between the ranges of expression in the tumor and the non-malignant controls; genes were excluded if >50% of the tumor samples showed expression levels within the 95th percentile of the non-malignant range. The variation in the degree of over-expression in the tumor samples reflects not only tumor heterogeneity but also variations in the extent of contamination of the tumor samples with "normal" tissue including muscle, stromal cells and non-malignant epithelial glands. This "normal" contamination ranged from 5 to 70% with a median of approximately 25%. Other genes were excluded because of high relative expression in hematopoietic cells, or elevated expression in metaplastic gastric tissue. It can be appreciated that depending on the degree of contamination by normal cells or cells that normally express the marker, different threshold ranges can be selected that can provide sufficient separation between a cancer source and a normal source.

Figure 2 depicts a table that shows results of studies using 34 markers for gastric malignancy selected using the above criteria. The table indicates the symbol for the gene, the MWG oligo number, the NCBI mRNA reference sequence number, the protein reference sequence number, the fold change between tumor and non-tumor gene expression, the fold change ranking relative to other cancer markers studied, the results of an original, unadjusted Student t-test, the results of the Bonferroni-adjusted p value and the results of the 2-sample Wilcoxon test.

The median fold change (tumor: non malignant tissue) for these 34 genes ranged from 1.6 to 7 and the median change in fold change rank ranged from –16,995 to –25,783. The maximum possible change in fold change rank was -29,718. For each of the markers shown, the statistical significance of their specificity as cancer markers was found to be extremely high. The Bonferroni-adjusted p values were, in general, all below 10^{-6} or less, indicating that diagnosis using these markers is very highly associated with gastric cancer.

The three cystatins (CST1, CST2, and CST4) are highly homologous and represented by the same oligonucleotide on the microarray.

All proteins depicted in Figure 2 were predicted to have signal peptides using the SMART package (European Molecular Biology Laboratory). Other proteins of this invention can also be released into extracellular fluid and have been detected in serum.

Each of the genes depicted in Figure 2 exhibited a change in intensity rank greater than the two oligonucleotides on the array corresponding to CEA, the marker most frequently used in clinical practice to monitor gastric cancer progression.

qPCR analysis

More sensitive and accurate quantitation of gene expression was obtained for a subset of the genes shown in Figure 3 using qPCR. RNA from 46 tumor and 49 non-malignant samples was analyzed for 23 genes identified by the microarray analysis (Figure 2) and results are shown in Figure 3. Figure 3 includes the gene symbol, median fold change between cancer and normal tissue, and the % T > 95^{th} percentile of expression levels in non-malignant samples. 12 tumor samples and 9 normal samples were excluded from the analysis because of high (>75%) normal cell contamination, a high degree of necrosis (>40%), or poor hybridisation signal on the microarrays. The median fold change (tumor tissue compared to non-malignant tissue) for these 23 genes ranged from 3 to 525 fold (Figure 3).

The level of expression of genes ASPN, CST1, 2 & 4, LOXL2, TIMP1, SPP1, SFRP4, INHBA, THBS2 and SPARC was greater in tumors than the 95^{th} percentile of the non-malignant range for ≥90% of cases (Figure 3). For the remainder of genes, the expression in tumors was greater than the 95^{th} percentile in >50% of samples. Each tumor over-expressed at least seven genes greater than the 95^{th} percentile indicating that combinations of markers will lead to comprehensive coverage of all gastric tumors (Figure 6).

Validation of Array Data

Array data was validated by carrying out using quantitative, real-time PCR (qPCR) on the tumor and non-malignant samples with probes for 24 genes. Of all 24 genes studied, 20 showed a strong correlation between the two techniques. Four of these analyses are show in Figures 4a – 4d, which depict graphs of the relative expression for

the 4 selected cancer markers detected using array and qPCR methods. For each graph in Figure 4, the horizontal axis represents the array log2 fold change in gene expression, and the vertical axis represents the qPCR log2 fold change in gene expression. We found that there was a strong correlation between the two methods, as indicated by the co-variant relationship between the methods. The strong correlation indicates that both microarray fold change analysis and qPCR are suitable methods for detecting changes in the expression of gastric cancer marker genes and therefore can be used as an accurate, sensitive screening method. It can also be appreciated from Figures 4a – 4d that qPCR can be more sensitive at detecting changes in expression than are array methods. Thus, in situations in which early detection is especially desirable, qPCR may be especially useful.

Figures 5a – 5w depict histograms comparing frequency of observation of expression of each of a series of 23 genes (vertical axis) and the log2 fold change in expression for that gene (horizontal axis), for both normal tissue (open bars) and tumor tissues (black bars). We found surprisingly that for each of these 23 genes, there was substantial separation in the frequency distributions between normal and tumor tissue, as reflected by the low degree of overlap between the frequency distribution curves. For example, Figure 5b depicts the results for CST 1, 2 and 4, for which there was only one normal sample observed to have an expression level in the tumor range. In other cases (e.g., Figure 5n; for PRS 11) each frequency distribution curve was relatively narrow and there was a degree of overlap. However, even for this marker, the median log2 fold change showed a substantial separation of the amount of gene expression. In other cases, (e.g., Figure 5a; ASPN), although there was some overlap, there was a clear separation of the median log2 fold expression between normal and tumor samples.

Figure 6 depicts a histogram of the number of genes exhibiting a significantly increased expression in tumor samples compared to normal samples (vertical axis) and the individual samples tested. In each case, the tumor sample exhibited multiple genes with elevated expression levels. The lowest number of genes having increased expression was 7, found in sample E123. This finding indicates that, in situations in which multiple genes are overexpressed relative to normal tissue, the reliability of cancer detection can be very high, making diagnosis of cancer more certain. However, in some

cases, elevation of expression of a single marker gene is sufficient to lead to the diagnosis of cancer.

Example 9: Detection of Gastric Tumor Marker Proteins

In yet further embodiments, GTM proteins can be detected as a basis for diagnosis. In certain situations, the concentration of mRNA in a particular sample, such as a sample containing no cells, it may be difficult to use either microarray or qPCR methods to detect elevations in gene expression. Thus, in certain embodiments, detection of GTM proteins can be accomplished using antibodies directed against either the entire protein, a fragment of the protein (peptide) or the protein core. Methods for detecting and quantifying expression of proteins and peptides are known in the art and can include methods relying on specific antibodies raised against the protein or peptide. Monoclonal antibodies and polyclonal antisera can be made using methods that are well known in the art and need not be described herein further.

Although certain marker proteins can be glycosylated, variations in the pattern of glycosylation can, in certain circumstances, lead to mis-detection of forms of GTMs that lack usual glycosylation patterns. Thus, in certain embodiments of this invention, GTM immunogens can include deglycosylated GTM or deglycosylated GTM fragments. Deglycosylation can be accomplished using one or more glycosidases known in the art. Alternatively, GTM cDNA can be expressed in glycosylation-deficient cell lines, such as prokaryotic cell lines, including E. coli, thereby producing non-glycosylated proteins or peptides. It can also be appreciated that the level and quality of glycosylation can be sensitive to the presence of essential precursors for sugar side-chains. Thus, in the absence of an essential sugar, "normal" glycosylation may not occur, but rather, shorter or missing side chain sugars may be found. Such "glycosylation variants" can be used as immunogens to produce antibodies specific for different types of marker genes.

Once an antibody or antiserum against a GTM is produced, such antibody preparations can be used for in a variety of ways. First, enzyme-linked immunosorbent assay (ELISA) or radioimmunoassay (RIA) methods can be used to quantify GTM proteins or peptides. Immunodetection can be accomplished in tissue samples using

immunohistochemistry. These methods are all known in the art and need not be described further herein.

Example 11: Vectors Containing GTM Oligonucleotides

Other embodiments of this invention include vectors useful for *in* vitro expression of marker genes or portions thereof ("marker peptides") or fragments of marker gene products. For example, vectors can be made having oligonucleotides for encoding GTMs therein. Many such vectors can be based on standard vectors known in the art. This invention also includes vectors that can be used to transfect a variety of cell lines to prepare GTM-producing cell lines, which can be used to produce desired quantities of GTMs for development of specific antibodies or other reagents for detection of GTMs or for standardizing developed assays for GTMs.

It is to be understood that to manufacture such vectors, an oligonucleotide containing the entire open reading frame or a portion of such an open reading frame encoding a portion of the protein to be expressed can be inserted into a vector containing a promoter region, one or more enhancer regions operably linked to the oligonucleotide sequece, with an initiation codon, an open reading frame, and a stop codon. Methods for producing expression vectors are known in the art and need not be repeated herein.

It can also be appreciated that one or more selectable markers can be inserted into an expression vector to permit the expansion of cell lines selected to contain the expression vector of interest. Moreover, one can also insert leader sequences known in the art, in frame, to direct secretion, internal storage or membrane insertion of the protein or protein fragment in the expressing cell.

Example 12: Cells Transfected with GTM-Containing Vectors

In still further embodiments, cells are provided that can express GTMs, GTM fragments or peptide markers. Both prokaryotic and eukaryotic cells can be so used. For example, E. coli (a prokaryotic cell) can be use to produce large quantities of GTMs lacking in mature glycosylation (if the particular GTM normally is glycosylated). COS cells, 293 cells and a variety of other eukaryotic cells can be used to produce GTMs that are glycosylated, or have proper folding and therefore, three-dimensional structure of the

native form of the GTM protein. Methods for transfecting such cells are known in the art and need not be described further herein.

Example 13: Kits

Based on the discoveries of this invention, several types of test kits can be produced. First, kits can be made that have a detection device pre-loaded with a detection molecule (or "capture reagent"). In embodiments for detection of GTM mRNA, such devices can comprise a substrate (e.g., glass, silicon, quartz, metal, etc) on which oligonucleotides as capture reagents that hybridize with the mRNA to be detected. In some embodiments, direct detection of mRNA can be accomplished by hybridizing mRNA (labeled with cy3, cy5, radiolabel or other label) to the oligonucleotides on the substrate. In other embodiments, detection of mRNA can be accomplished by first making complementary DNA (cDNA) to the desired mRNA. Then, labeled cDNA can be hybridized to the oligonucleotides on the substrate and detected.

Regardless of the detection method employed, comparison of test GTM expression with a standard measure of expression is desirable. For example, RNA expression can be standardized to total cellular DNA, to expression of constitutively expressed RNAs (for example, ribosomal RNA) or to other relatively constant markers.

Antibodies can also be used in kits as capture reagents. In some embodiments, a substrate (e.g., a multiwell plate) can have a specific GTM capture reagent attached thereto. In some embodiments, a kit can have a blocking reagent included. Blocking reagents can be used to reduce non-specific binding. For example, non-specific oligonucleotide binding can reduced using excess DNA from any convenient source that does not contain GTM oligonucleotides, such as salmon sperm DNA. Non-specific antibody binding can be reduced using an excess of a blocking protein such as serum albumin. It can be appreciated that numerous methods for detecting oligonucleotides and proteins are known in the art, and any strategy that can specifically detect GTM associated molecules can be used and be considered within the scope of this invention.

In embodiments relying upon antibody detection, GTM proteins or peptides can be expressed on a per cell basis, or on the basis of total cellular, tissue, or fluid protein,

fluid volume, tissue mass (weight). Additionally, GTM in serum can be expressed on the basis of a relatively high-abundance serum protein such as albumin.

In addition to a substrate, a test kit can comprise capture reagents (such as probes), washing solutions (e.g., SSC, other salts, buffers, detergents and the like), as well as detection moieties (e.g., cy3, cy5, radiolabels, and the like). Kits can also include instructions for use and a package.

Although this invention is described with reference to specific embodiments thereof, it can be appreciated that other embodiments involving the use of the disclosed markers can be used without departing from the scope of this invention.

INDUSTRIAL APPLICABILITY

Methods for detecting GTM family members include detection of nucleic acids using microarray and/or real time PCR methods and detection of proteins and peptides. The compositions and methods of this invention are useful in the manufacture of diagnostic devices and kits, diagnosis of disease, evaluating efficacy of therapy, and for producing reagents suitable for measuring expression of GTM family members in biological samples.

I Claim:

1. A method for detecting cancer, comprising:
   (a) providing a biological sample; and
   (b) detecting the over expression of a GTM family member in said sample.

2. The method of claim 1, wherein said cancer is gastric cancer.

3. The method of claim 1, wherein said step of detecting is carried out by detecting over expression of GTM mRNA.

4. The method of claim 1, wherein said step of detecting is carried out by detecting over expression of a GTM protein.

5. The method of claim 1, wherein said step of detecting is carried out by detecting over expression of a GTM peptide.

6. A polyclonal antiserum specific for a GTM.

7. A monoclonal antibody specific for a GTM.

8. A device for detecting a GTM, comprising:
   a substrate having a GTM capture reagent thereon; and
   a detector associated with said substrate, said detector capable of detecting a GTM associated with said capture reagent.

9. The device of claim 8, wherein said GTM capture reagent is an oligonucleotide.

10. The device of claim 8, wherein said GTM capture reagent is an antibody.

11. A kit for detecting cancer, comprising:

a substrate;
a GTM capture reagent; and
instructions for use.

12. The kit of claim 11, wherein said GTM capture reagent is a GTM-specific oligonucleotide.

13. The kit of claim 11, wherein said GTM capture reagent is a GTM-specific antibody.

14. An expression vector for a GTM, comprising:
a promoter;
one or more enhancers;
an initiation codon;
an oligonucleotide sequence of a GTM in an open reading frame; and
a stop codon.

15. The vector of claim 14, further comprising a selectable marker.

16. A cell for expressing a GTM comprising:
a cell capable of sustained growth in *in vitro* conditions; and
a functional expression vector containing a GTM oligonucleotide sequence.

17. The cell of claim 16, wherein said cell is a prokaryotic cell.

18. The cell of claim 16, wherein said cell is a eukaryotic cell.

19. A method for manufacturing a monoclonal antibody, comprising the steps of:
immunizing a host organism with a GTM protein or fragment;
isolating a spleen cell from said host;

fusing said spleen cell with a cell capable of being propagated *in vitro* thereby producing a fused cell;

selecting and isolating said fused cell; and

producing an culture from said isolated fused cells.

20. A method for detecting gastric cancer, comprising the steps of:

providing a sample from a patient suspected of having gastric cancer;

measuring the presence of a GTM protein using an ELISA method.

ABSTRACT

Early detection of tumors is a major determinant of survival of patients suffering from tumors, including gastric tumors. Members of the GTM gene family can be highly and consistently over-expressed in gastric tumor tissue and other tumor tissue, and thus can be used as markers for gastric and other types of cancer. GTM proteins can be secreted and can reach high concentrations in the serum and/or other fluids when expressed, detection of GTM family members can be also be detected in the body fluids of individuals with cancer and thus can be an effective diagnostic approach. Anti-GTM antibodies are useful markers for the diagnosis, early detection and monitoring of gastric and other types of cancer.

name	symbol	Applied Biosystems "assay on demand" assay #	forward primer	Seq ID No.	reverse primer	Seq. ID No.	probe	Seq. ID No.
asporin (lrr class 1)	ASPN		AAATACAAAAGGACACATTCAAAGG	1	TGCTTCTGCAATTCTGATATGGA	18	TTGGAAATGAGTGCAAACCCTCTTGATAATAATG	35
chondroitin sulfate proteoglycan 2 (versican)	CSPG2		GCCAGTGGAATGATGTTCCC	2	TCTTGGCATTTTCTACAACAGGG	19	AGGAACAGTTGCTTGCGGCCAGC	36
cystatins SN, SA & S	CST1, 2, 4		AGTCCCAGCCCAACTTGGA	3	GGGAACTTCGTAGATCTGGAAAGA	20	AGCCAGAACTGCAGAAGAAACAGTTGTGC	37
gamma-glutamyl hydrolase	GGH		GTGGCAATGCCGCTGAA	4	TGACAGCAACAACTCAGTAGGAAAA	21	TTCACTGGAGGTCAATTGCACAGCAGAAT	38
insulin-like growth factor binding protein 7	IGFBP7		CAGGTCAGCAAGGGCACC	5	TCACAGCTCAAGTACACCTGGG	22	AGCAAGGTCCTTCCATAGTGACGCCC	39
kallikrein 10	KLK10		ACAACATGATATGTGCTGGACTGG	6	GAGAGGATGCCTTGGAGGGT	23	CTTGCCAGAGTGACTCTGGAGGCCC	40
leucine proline-enriched proteoglycan 1(leprecan 1)	LEPRE1		CTTGAGTACAACGCTGACCTCTTC	7	CCGTGACACAGTTCTGCTTACAG	24	CCATCACAGATCATTACATCCAGGTCCTCA	41
lumican	LUM		GATTCTTGTCCATAGTGCATCTGC	8	CCAATCAATGCCAGGAAGAGA	25	TAAGGATTCAAACCATTTGCCAAAAATGAGTCTAAG	42
lysyl oxidase-like 2	LOXL2		AGGCCAGCTTCTGCTTGGA	9	CCCTGATCGCCGAGTTG	26	CGTAATTCTTCTGGATGTCTCCTTCACATTCTG	43
matrix metalloproteinase 12	MMP12		GCCTCTCTGCTGATGACATACGT	10	AGTGACAGCATCAAAACTCAAATTG	27	TCAGTCCCTGTATGGAGACCCAAAAGAGAA	44
metalloproteinase inhibitor 1	TIMP1		CCAGACCACCTTATACCAGCG	11	GGACCTGTGGAAGTATCCGC	28	CAAGATGACCAAGATGTATAAAGGGTTCCAAGC	45
n-acylsphingosine amidohydrolase	ASAH1		CGCAGAACGCCTGCAAA	12	ACAGGACATCATACATGGTTTCAAA	29	TGTCTGAACCGCACCAGCCAAGAGAATA	46
secreted frizzled-related protein 2	SFRP2		CGCTAGCAGCGACCACCT	13	TTTTGCAGGCTTCACATACCTTT	30	CTGCCAGCCACCGAGGAAGCTC	47
secreted protein, acidic, cysteine rich	SPARC		TCTTCCCTGTACACTGGCAGTTC	14	GAAAAAGCGGGTGGTGCA	31	TGGACCAGCACCCCATTGACGG	48
serine protease 11 (IGF binding)	PRSS11		TCGGGAGGCCCGTTAGTAA	15	AAGGAGATTCCAGCTGTCACTTTC	32	AGTGTTAATTCCAATCACTTCACCGTCCAGG	49
thrombospondin 2	THBS2		TGGAAGGACTACACGGCCTATAG	16	TAGGTTTGGTCATAGATAGGTCCTGAG	33	AGGCCCAAGACCGGCTACATCAGAGTC	50
thyroglobulin	TG		GACGGTTCCTCGCAGTTCAA	17	TGTAAACCGCTCCACTTCACAT	34	TCTGGCAGATTCCGATGCCCCACAA	51
adlican	-	Hs00377849_m1						
egf-containing fibulin-like extracellular matrix protein 2	EFEMP2	Hs00213545_m1						
secreted frizzled-related protein 4	SFRP4	Hs00180066_m1						
inhibin beta A chain	INHBA	Hs00170103_m1						
osteopontin	SPP1	Hs00167093_m1						
transforming growth factor B-induced	TGFBI	Hs00165908_m1						

Figure 1

Microarray - Identification of Markers for Gastric Malignancy

name	symbol	MWG oligo #	NCBI mRNA ref sequence	protein ref sequence	fold change	fold change rank	original t-test	Bonferroni-adjusted p value	2 sample Wilcoxon test
adlican	-	C:0531	NM_015419	NP_056234	1.8	-17818	1.0E-28	3.04E-24	0.0E+00
asporin (lrr class 1)	ASPN	A:07749	NM_017680	NP_060150	2.6	-22292	6.4E-23	1.9E-18	0.0E+00
carboxypeptidase N	CPN2	B:4922	-	P22792	2.7	-22367.5	2.3E-42	7.0E-38	0.0E+00
cell growth regulatory factor with EF-hand domain	CGR11	A:07876	NM_006569	NP_006560	3.0	-21188.5	4.33E-42	1.3E-37	0.0E+00
chondroitin sulfate proteoglycan 2 (versican)	CSPG2	A:10008	NM_004385	NP_004376	2.3	-21606.5	2.23E-33	6.65E-29	0.00E+00
cystatin SN	CST1	A:06089	NM_001898	NP_001889	2.1	-17475	1.3E-18	3.8E-14	0.0E+00
cystatin SA	CST2	A:06089	NM_001322	NP_001313	2.1	-17475	1.3E-18	3.8E-14	0.0E+00
cystatin S	CST4	A:06089	NM_001899	NP_001890	2.1	-17475	1.3E-18	3.8E-14	0.0E+00
egf-containing fibulin-like extracellular matrix protein 2	EFEMP2	A:09072	NM_016938	NP_058634	2.4	-22761	2.0E-35	5.9E-31	0.0E+00
gamma-glutamyl hydrolase	GGH	A:03601	NM_003878	NP_003869	1.6	-18092	1.6E-07	4.8E-03	5.7E-11
inhibin beta A chain	INHBA	A:02189	NM_002192	NP_002183	2.1	-21247	1.4E-30	4.3E-26	0.0E+00
insulin-like growth factor binding protein 7	IGFBP7	A:03385	NM_001553	NP_001544	3.0	-25854	5.4E-31	1.6E-26	0.0E+00
kallikrein 10	KLK10	A:07907	NM_002776	NP_002767	2.3	-17986.5	5.0E-10	1.5E-05	4.9E-06
leucine proline-enriched proteoglycan 1(leprecan 1)	LEPRE1	A:04646	NM_022356	NP_071751	1.7	-18019	8.2E-14	2.4E-09	1.1E-12
lumican	LUM	A:09199	NM_002345	NP_002336	2.9	-24927	4.2E-24	1.3E-19	0.0E+00
lysyl oxidase-like 2	LOXL2	A:06085	NM_002318	NP_002309	1.6	-16994.5	5.9E-10	1.7E-05	7.9E-10
matrix metalloproteinase 12	MMP12	A:01762	NM_002426	NP_002417	2.1	-20209.5	2.2E-12	6.6E-08	4.9E-11
metalloproteinase inhibitor 1	TIMP1	A:08048	NM_003254	NP_003245	3.2	-24177	7.5E-38	2.3E-33	0.0E+00
n-acylsphingosine amidohydrolase	ASAH1	A:10030	NM_004315	NP_004306	1.7	-19636.5	9.6E-16	2.9E-11	0.0E+00
olfactomedin	OLFM1	B:3555	NM_014279	NP_055094	3.9	-25782.5	6.5E-46	1.9E-41	0.0E+00
osteopontin	SPP1	A:09441	NM_000582	NP_000573	7.0	-26668	4.0E-32	1.2E-27	0.0E+00
group xiii secreted phospholipase a2	PLA2G13	B:1811	NM_032562	NP_115951	3.0	-23212	7.92E-39	2.36E-34	0.00E+00
secreted frizzled-related protein 2	SFRP2	B:1634	XM_050625	XP_050625	2.1	-19217	2.7E-10	8.1E-06	4.1E-08
secreted frizzled-related protein 4	SFRP4	A:07398	NM_003014	NP_003005	3.0	-22153	6.0E-24	1.8E-19	0.0E+00
serine (or cysteine) proteinase inhibitor	SERPINH1	A:08615	NM_001235	NP_001226	1.9	-20252	2.8E-34	8.2E-30	0.0E+00
serine protease 11 (IGF binding)	PRSS11	B:1274	NM_002775	NP_002766	1.6	-17184.5	9.3E-18	2.8E-13	0.0E+00
secreted protein, acidic, cysteine rich	SPARC	A:08092	NM_003118	NP_003109	2.5	-22947.5	1.5E-44	4.6E-40	0.0E+00
spondin 2	SPON2	B:2543	NM_012445	NP_036577	2.4	-20390.5	2.9E-31	8.5E-27	0.0E+00
stannin	SNN	A:09316	NM_003498	NP_003489	2.1	-20162.5	3.25E-24	9.71E-20	0.00E+00
thrombospondin 2	THBS2	B:9017	NM_003247	NP_003238	2.6	-22095	5.8E-29	1.7E-24	0.0E+00
thrombospondin repeat containing 1	TSRC1	B:7686	NM_019032	NP_061905	2.6	-22608	1.38E-45	4.1E-41	0.0E+00
thyroglobulin	TG	B:5402	NM_003235	NP_003226	2.4	-23644	4.39E-36	1.3E-31	0.0E+00
transforming growth factor B-induced	TGFBI	A:08124	NM_000358	NP_000349	2.5	-23339.5	1.96E-24	9.71E-20	0.0E+00
transmembrane 6 superfamily member 2	TM6SF2	C:6300	NM_023002	NP_075378	3.4	-23516.5	7.32E-44	2.2E-39	0.0E+00

Figure 2

Quantitative RT-PCR - Quantification of Expression of Selected Gastric Cancer Candidate Genes

name	symbol	median T:N fold change	Maximum T:N fold change	% T >95th percentile [1]
adlican		5	37	74
asporin (lrr class 1)	ASPN	12	73	91
chondroitin sulfate proteoglycan 2 (versican)	CSPG2	6	24	78
cystatins SN, SA & S	CST1, 2, 4	525	25532	100
egf-containing fibulin-like extracellular matrix protein 2	EFEMP2	3	15	56
gamma-glutamyl hydrolase	GGH	5	36	67
inhibin beta A chain	INHBA	34	357	98
insulin-like growth factor binding protein 7	IGFBP7	4	19	80
kallikrein 10	KLK10	5	633	70
leucine proline-enriched proteoglycan 1(leprecan 1)	LEPRE1	4	17	72
lumican	LUM	5	47	80
lysyl oxidase-like 2	LOXL2	6	26	93
matrix metalloproteinase 12	MMP12	9	586	67
metalloproteinase inhibitor 1	TIMP1	8	19	91
n-acylsphingosine amidohydrolase	ASAH1	3	7	63
osteopontin	SPP1	40	481	96
secreted frizzled-related protein 2	SFRP2	5	85	63
secreted frizzled-related protein 4	SFRP4	56	600	100
secreted protein, acidic, cysteine rich	SPARC	9	56	93
serine protease 11 (IGF binding)	PRSS11	4	25	54
thrombospondin 2	THBS2	25	239	91
thyroglobulin	TG	5	153	54
transforming growth factor B-induced	TGFBI	7	204	82

[1] percentage of tumors with expression levels greater than the 95th percentile of non-malignant samples.

Figure 3

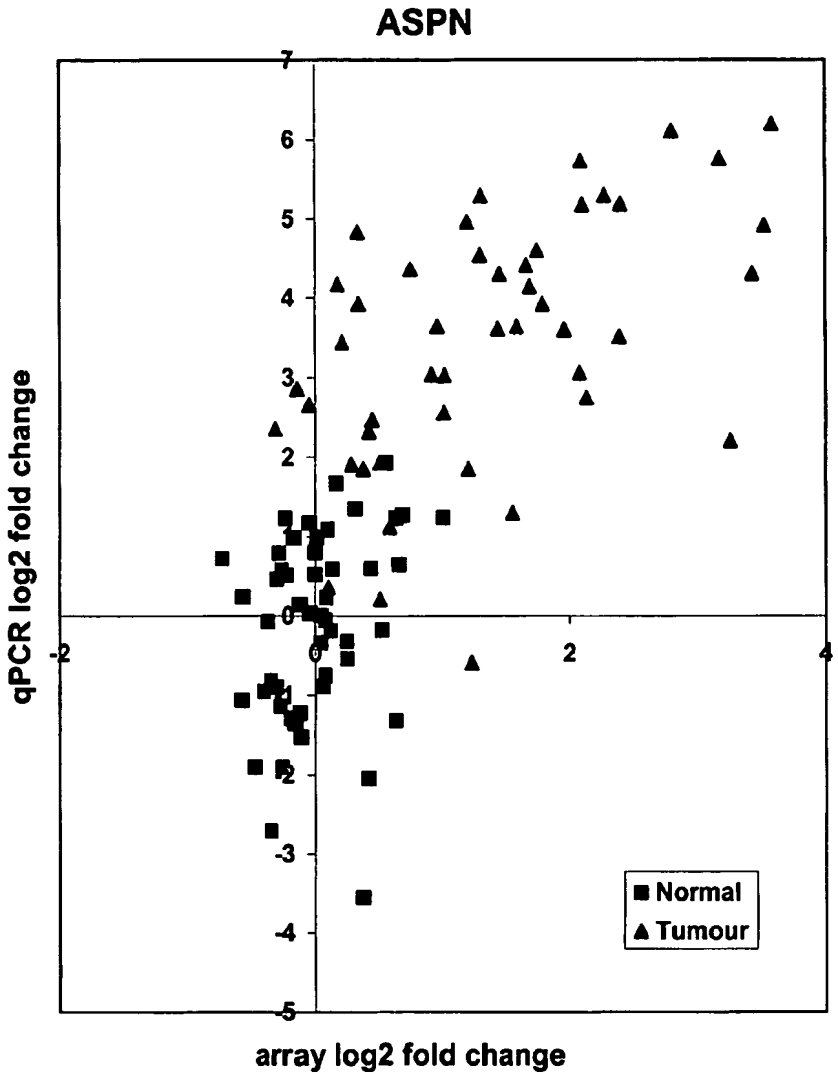
ASPN
qPCR log2 fold change
array log2 fold change
Normal
Tumour

Figure 4(a)

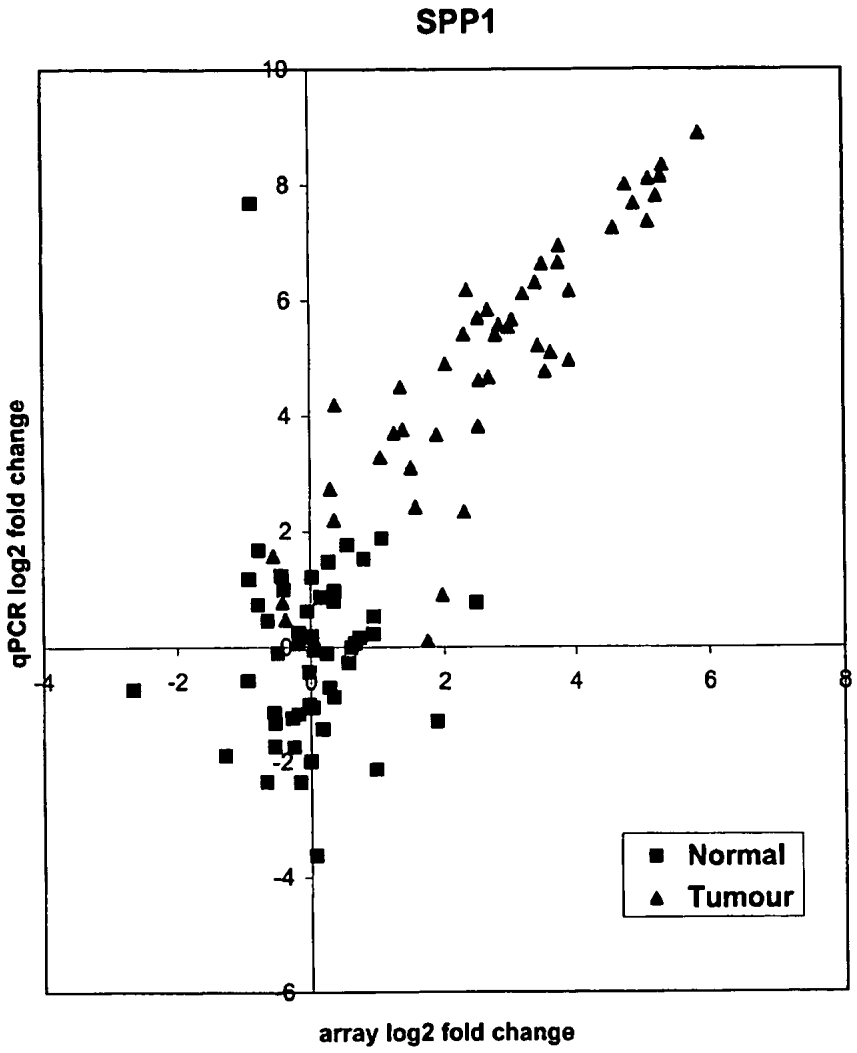
SPP1
qPCR log2 fold change
array log2 fold change
10
8
6
4
2
0
-2
-4
-6
-4
-2
0
2
4
6
8
Normal
Tumour

Figure 4(b)

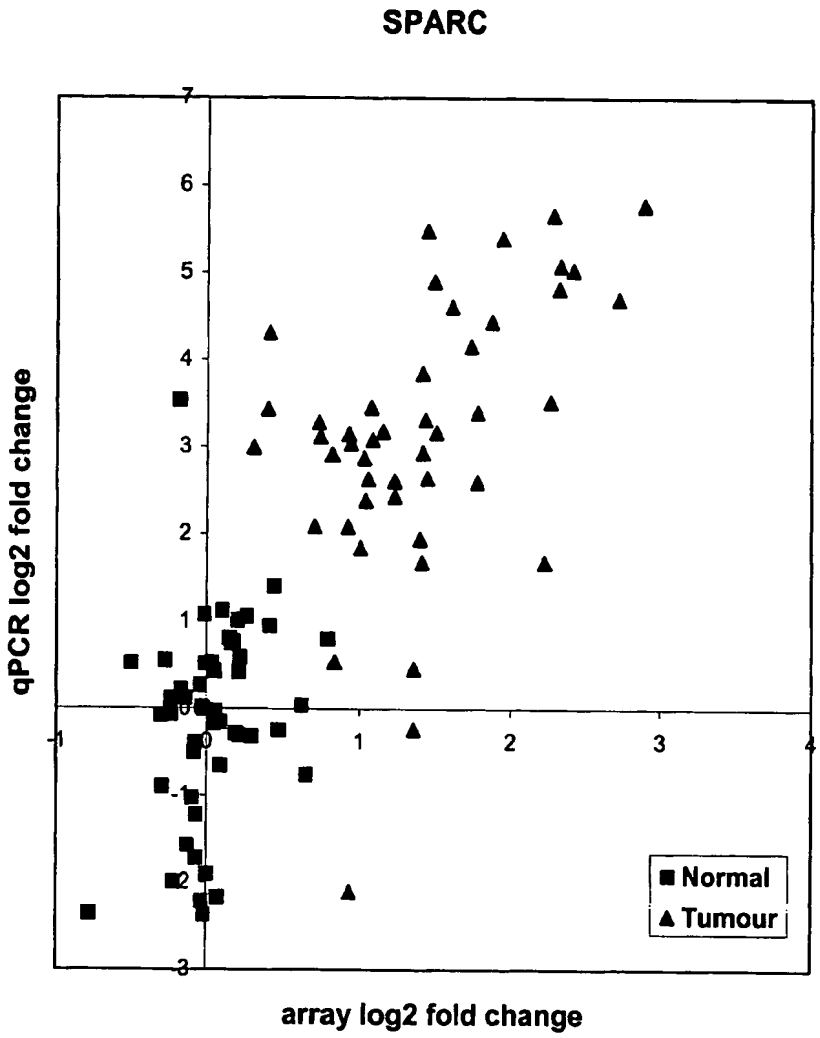
SPARC
qPCR log2 fold change
array log2 fold change
Normal
Tumour
7
6
5
4
3
2
1
0
-1
-2
-3
-1
0
1
2
3
4

Figure 4(c)

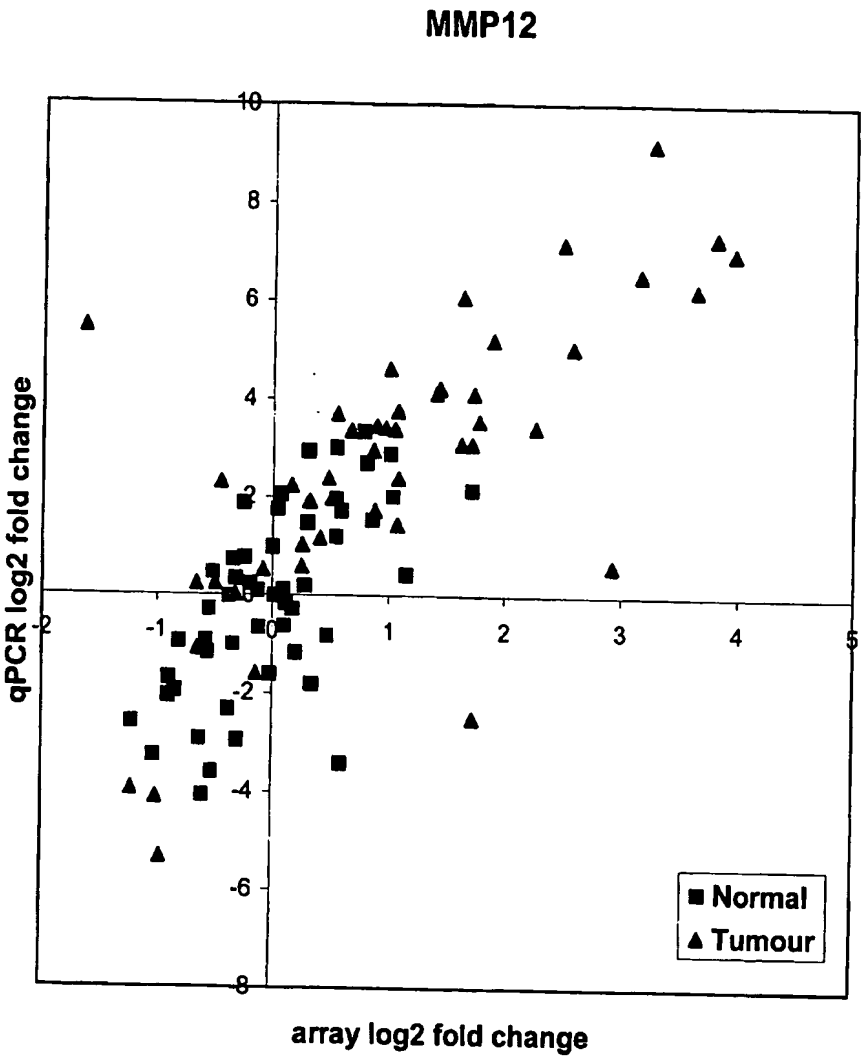
MMP12
qPCR log2 fold change
array log2 fold change
Normal
Tumour
10
8
6
4
2
-2
-4
-6
-8
-2
-1
0
1
2
3
4
5

Figure 4(d)

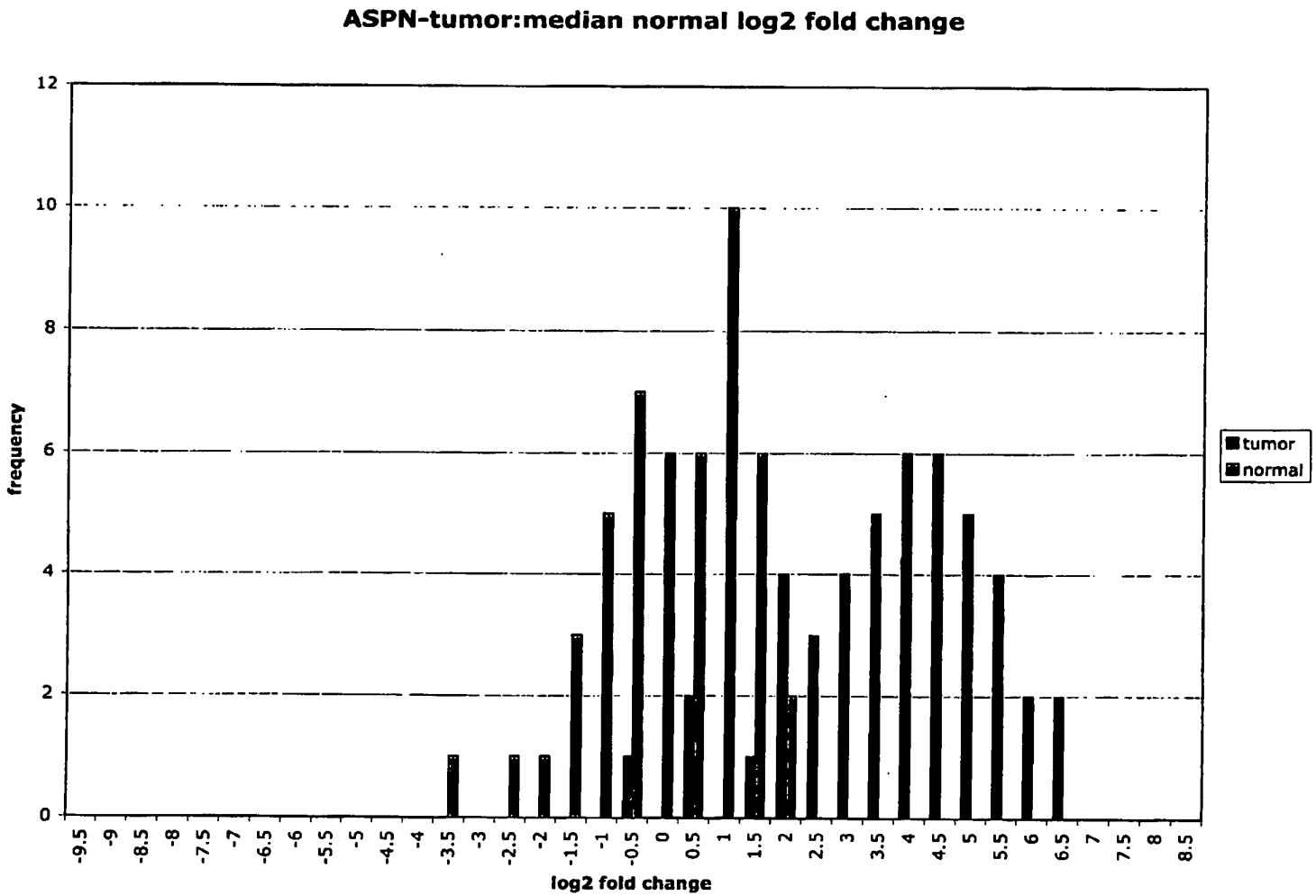
ASPN-tumor:median normal log2 fold change
frequency
log2 fold change
tumor
normal

Figure 5(a)

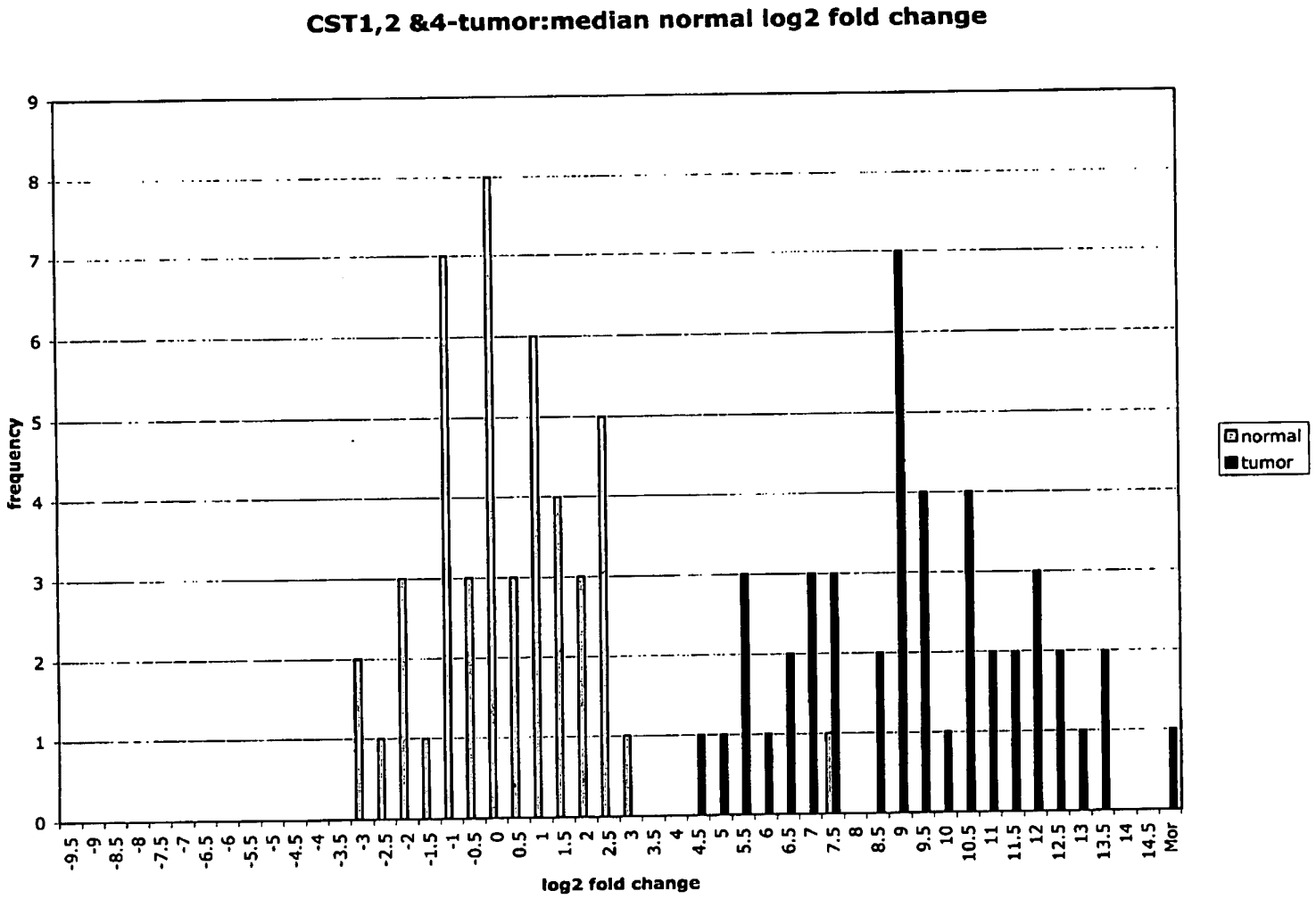
CST1,2 &4-tumor:median normal log2 fold change
frequency
log2 fold change
normal
tumor

Figure 5(b)

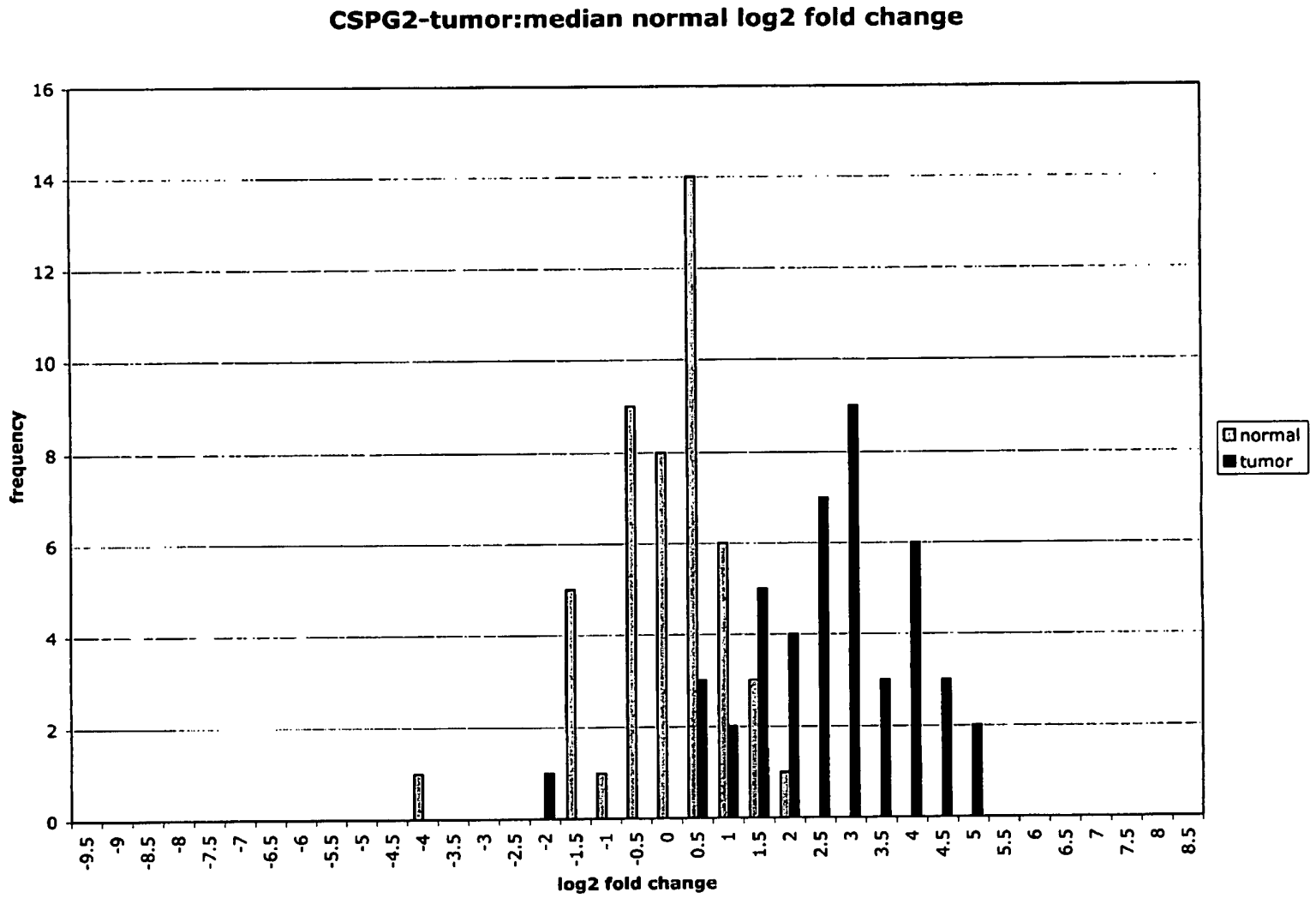
CSPG2-tumor:median normal log2 fold change
frequency
log2 fold change
normal
tumor

Figure 5(c)

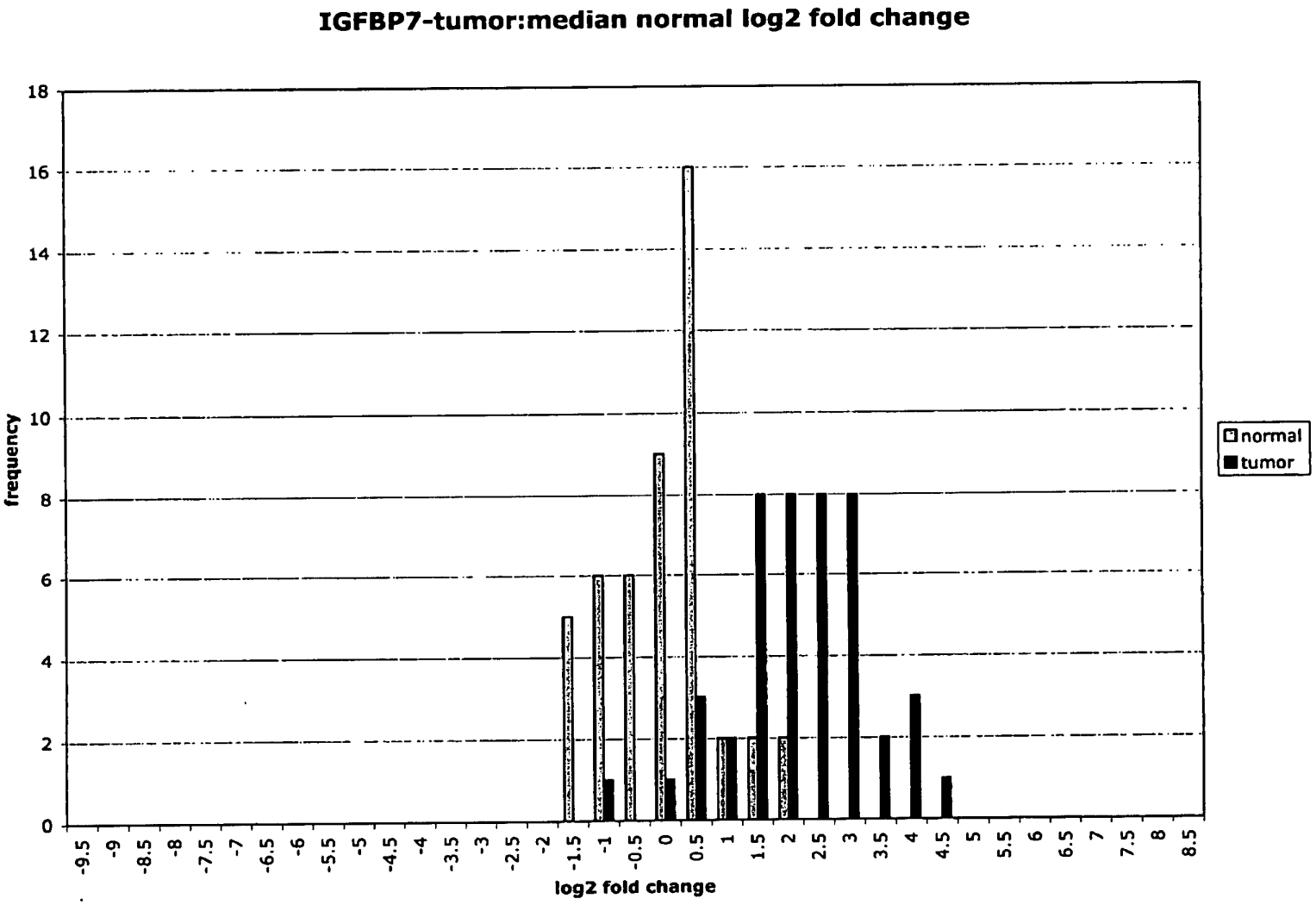
IGFBP7-tumor:median normal log2 fold change
frequency
log2 fold change
normal
tumor

Figure 5(d)

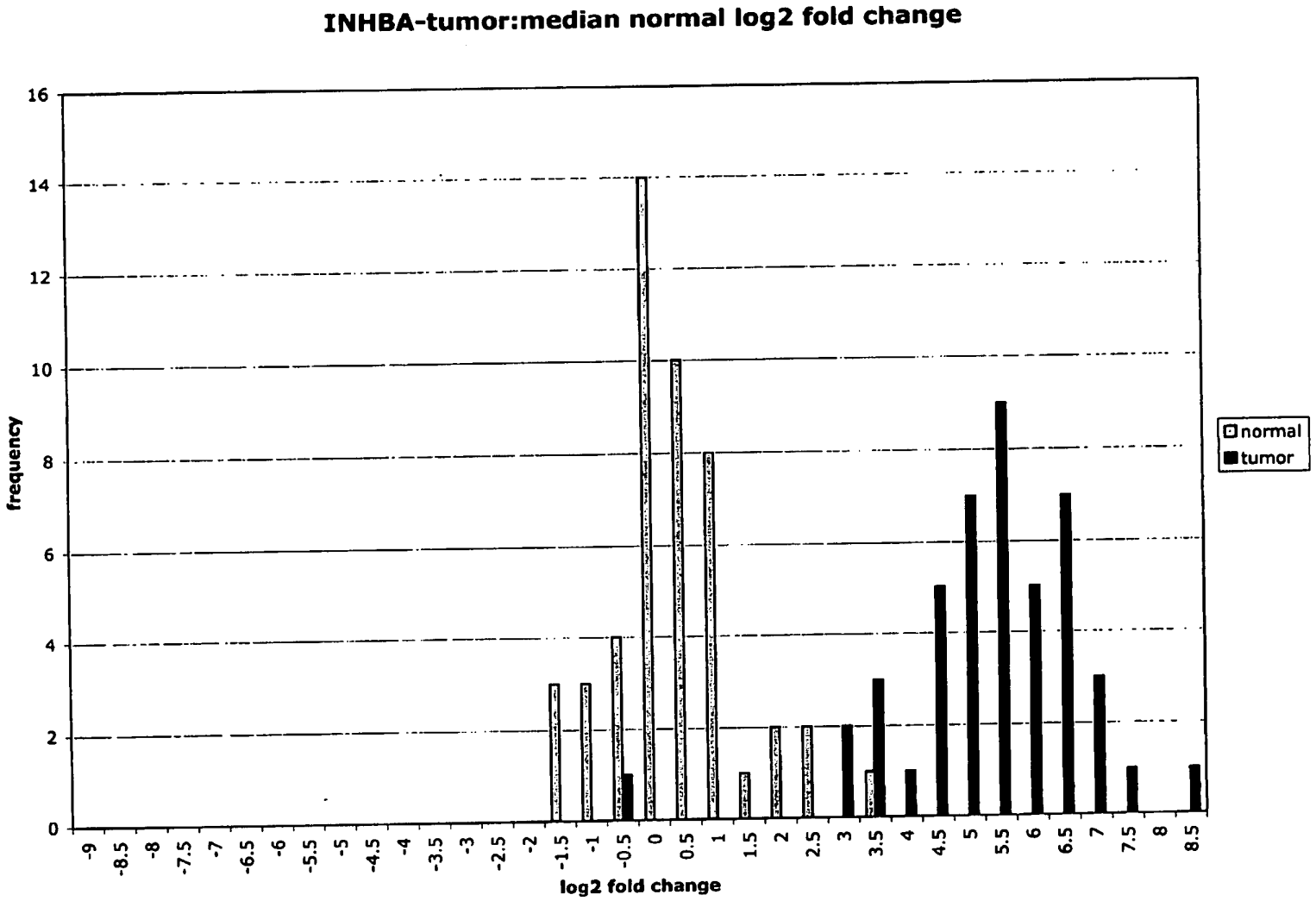
INHBA-tumor:median normal log2 fold change
frequency
log2 fold change
normal
tumor
0
2
4
6
8
10
12
14
16
-9
-8.5
-8
-7.5
-7
-6.5
-6
-5.5
-5
-4.5
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
8
8.5

Figure 5(e)

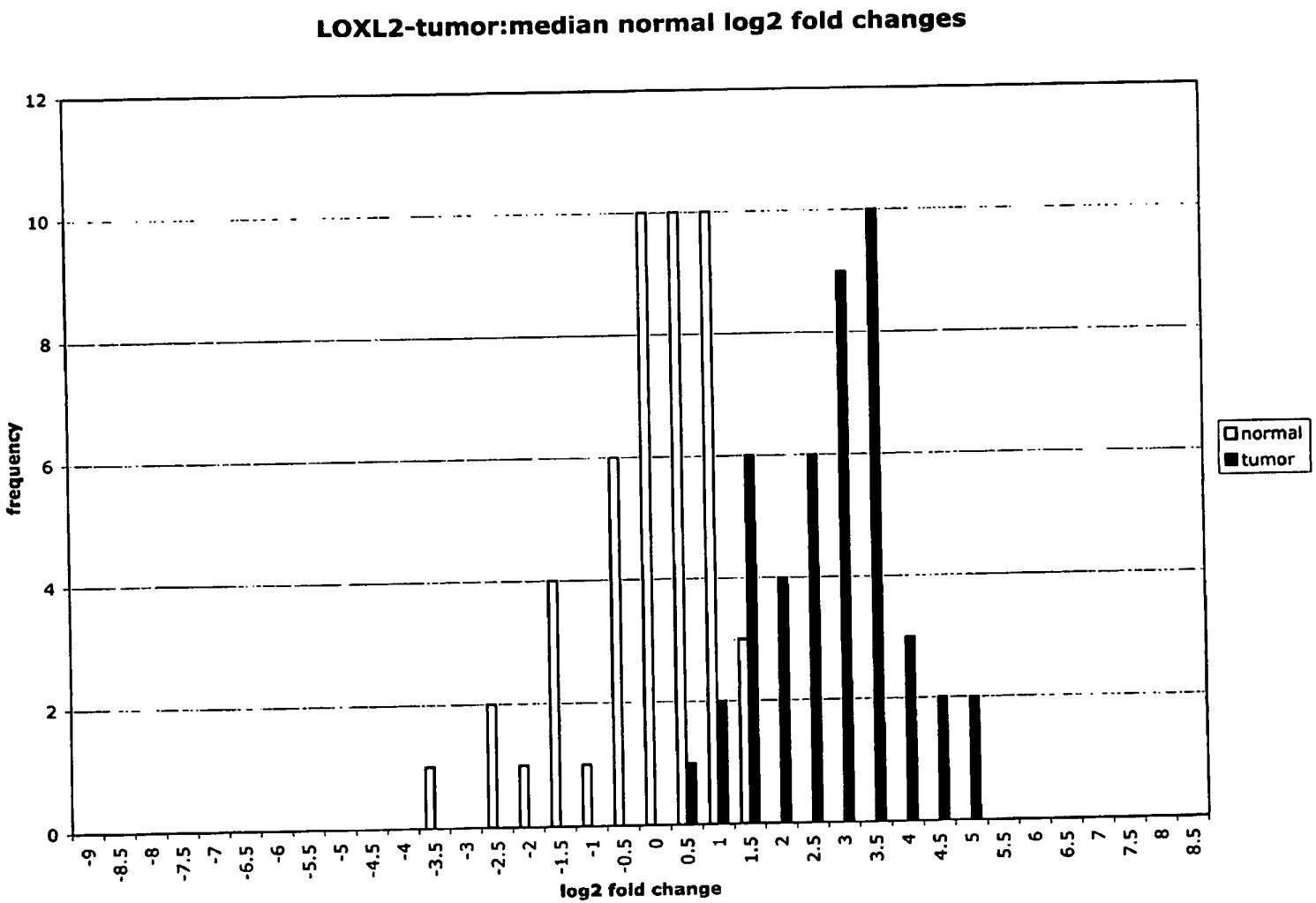
LOXL2-tumor:median normal log2 fold changes
frequency
log2 fold change
normal
tumor

Figure 5(f)

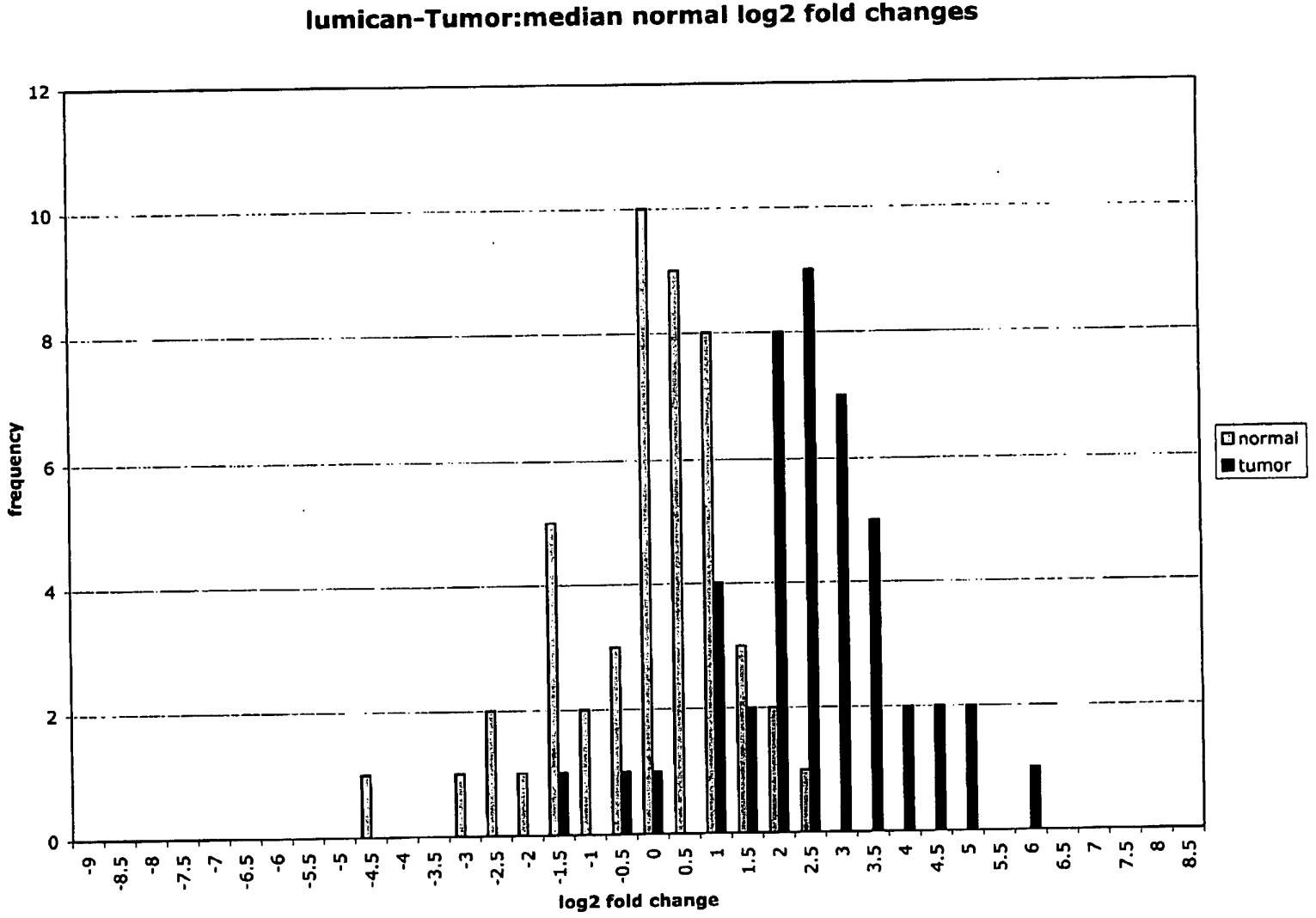
lumican-Tumor:median normal log2 fold changes
frequency
log2 fold change
normal
tumor

Figure 5(g)

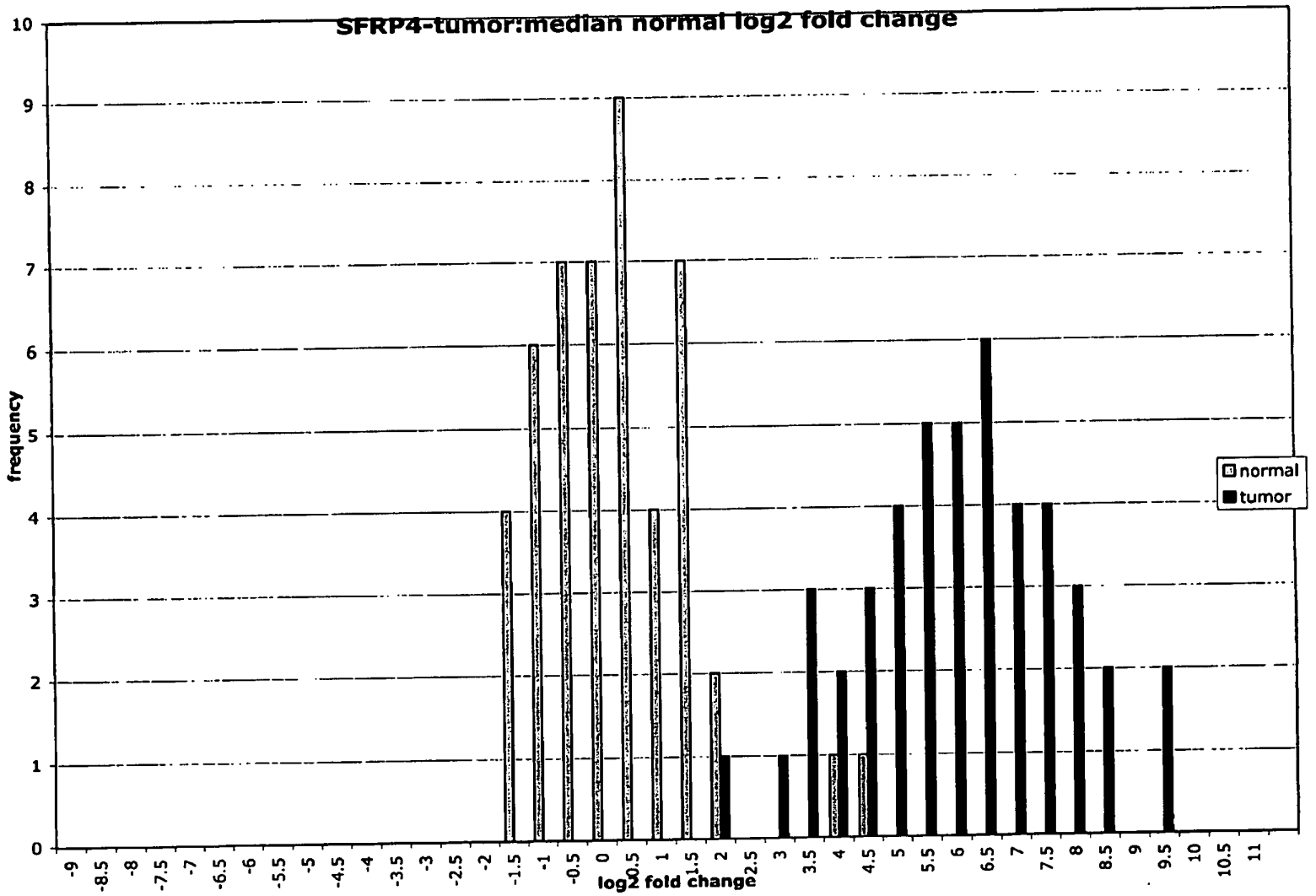
SFRP4-tumor:median normal log2 fold change
frequency
log2 fold change
normal
tumor

Figure 5(h)

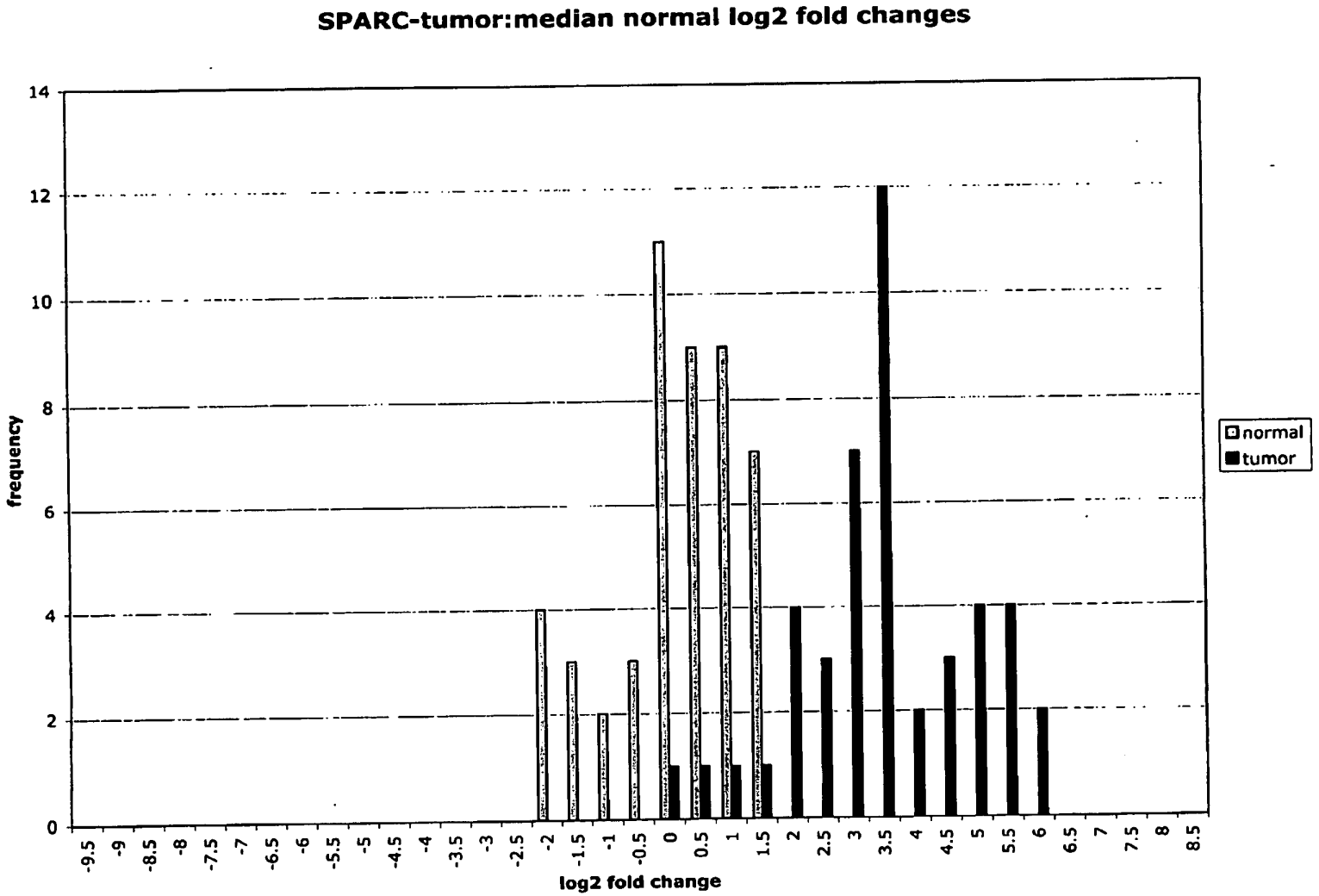
SPARC-tumor:median normal log2 fold changes
frequency
log2 fold change
normal
tumor
0
2
4
6
8
10
12
14
-9.5
-9
-8.5
-8
-7.5
-7
-6.5
-6
-5.5
-5
-4.5
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
8
8.5

Figure 5(i)

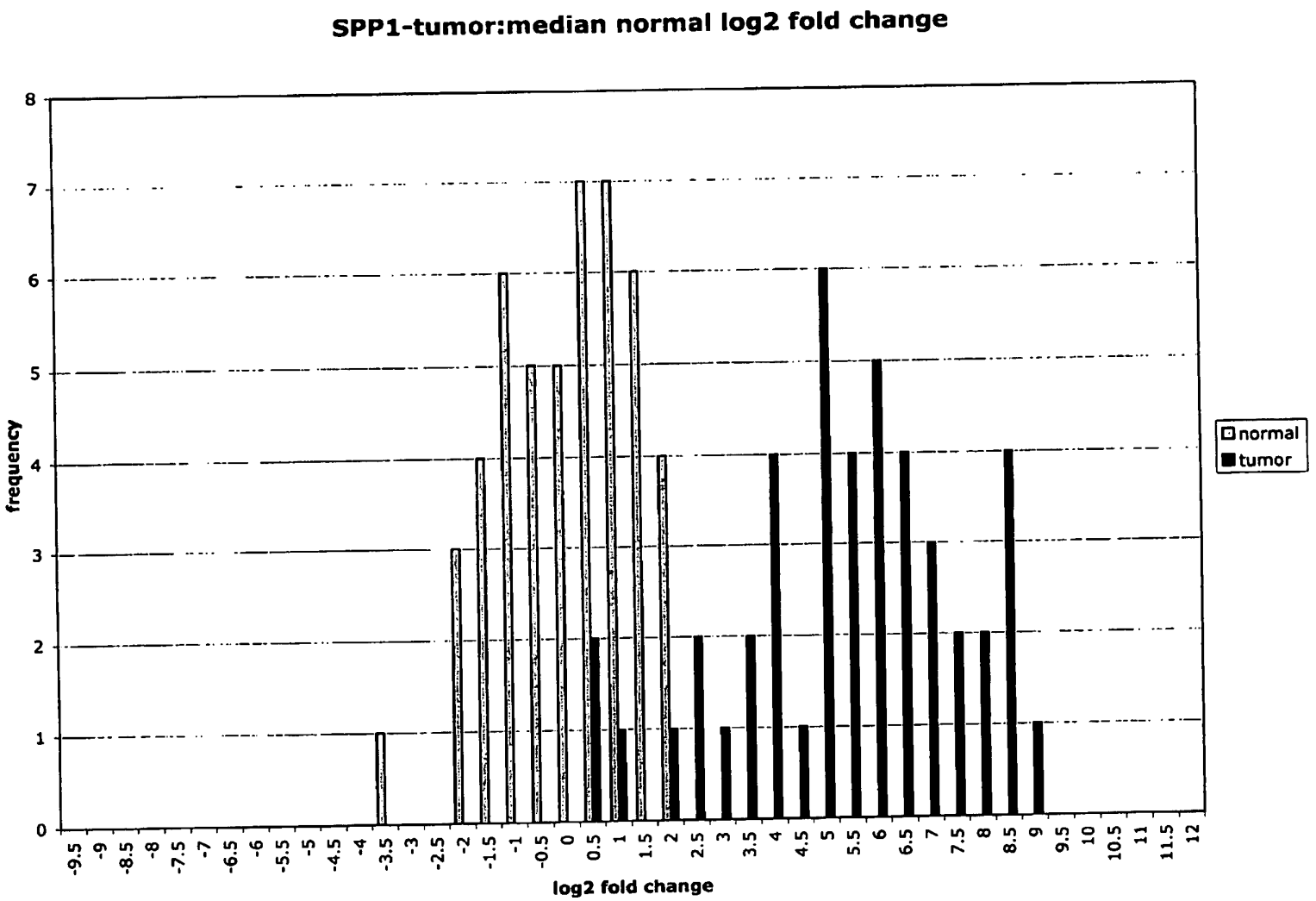
SPP1-tumor:median normal log2 fold change
frequency
log2 fold change
normal
tumor

Figure 5(j)

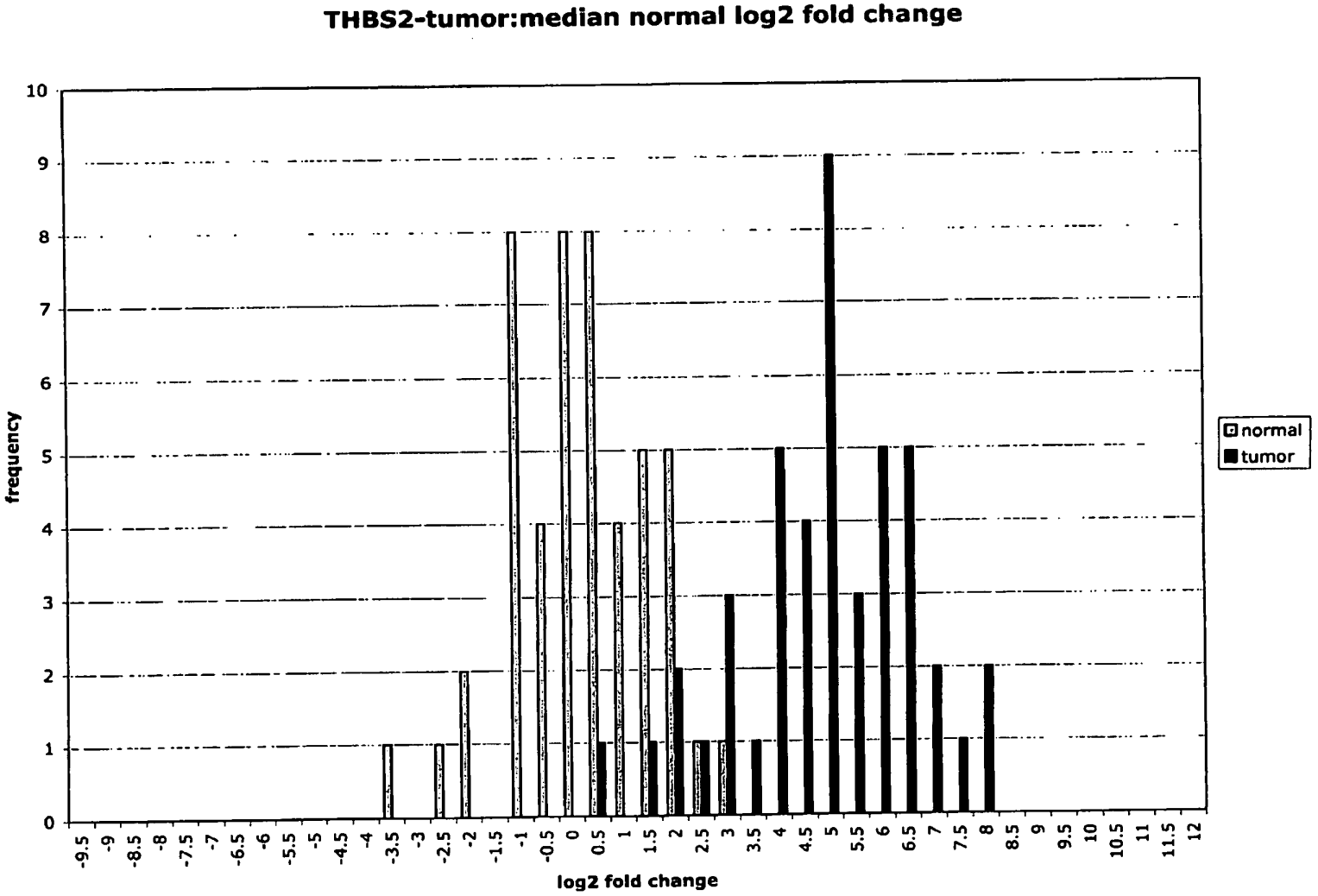
THBS2-tumor:median normal log2 fold change
frequency
log2 fold change
normal
tumor

Figure 5(k)

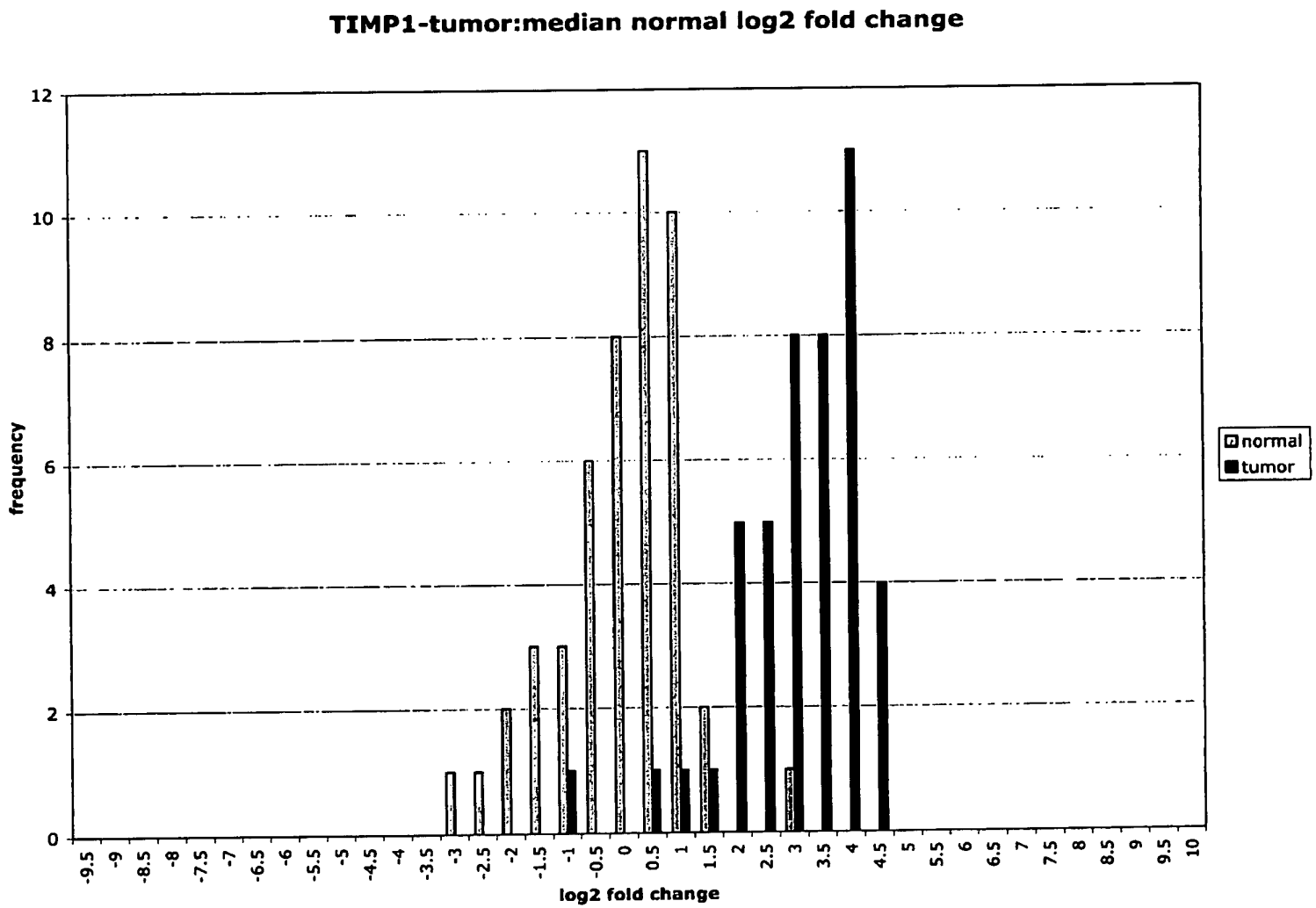
TIMP1-tumor:median normal log2 fold change
frequency
log2 fold change
normal
tumor

Figure 5(l)

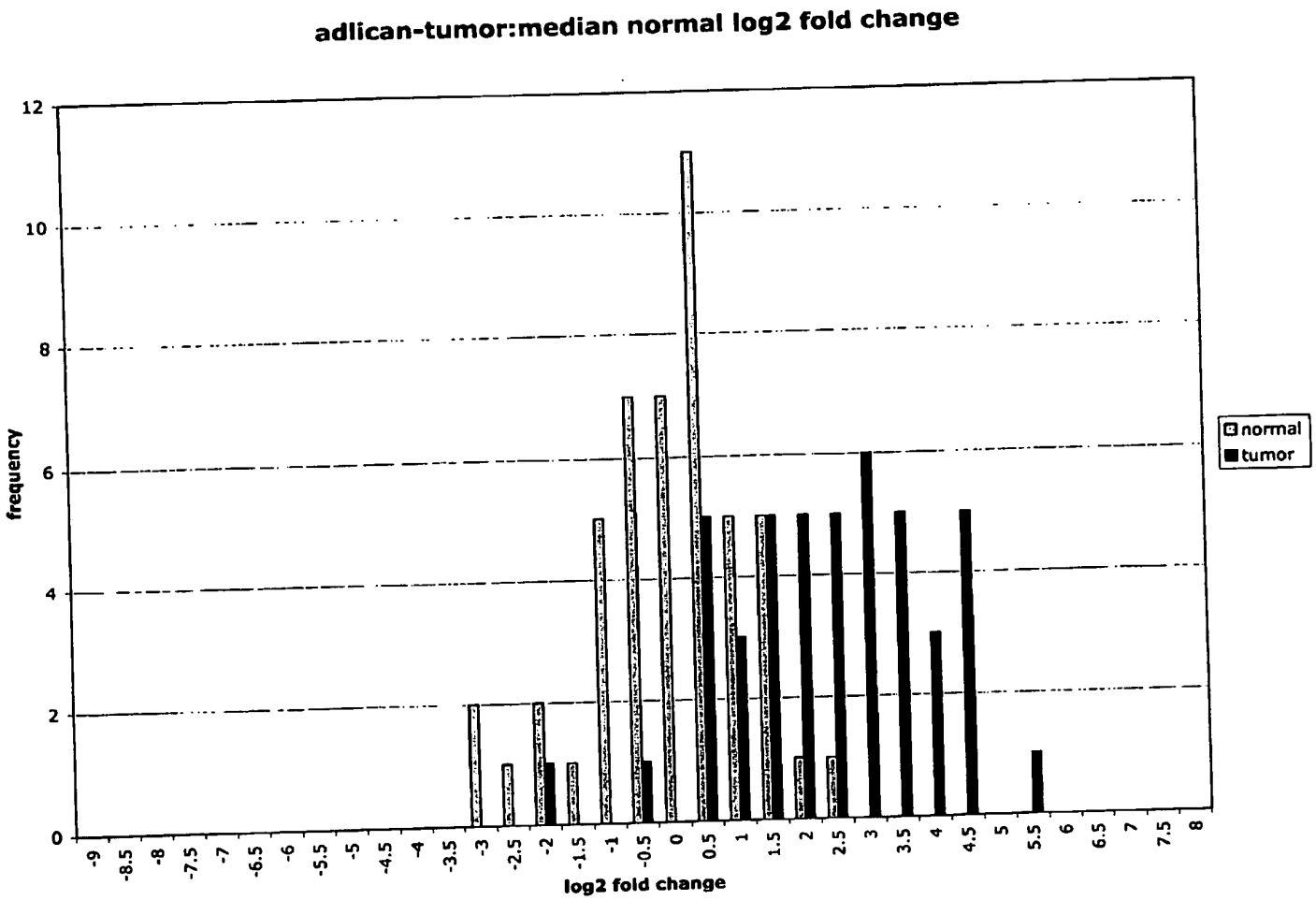
adlican-tumor:median normal log2 fold change
frequency
log2 fold change
normal
tumor
0
2
4
6
8
10
12
-9
-8.5
-8
-7.5
-7
-6.5
-6
-5.5
-5
-4.5
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
8

Figure 5(m)

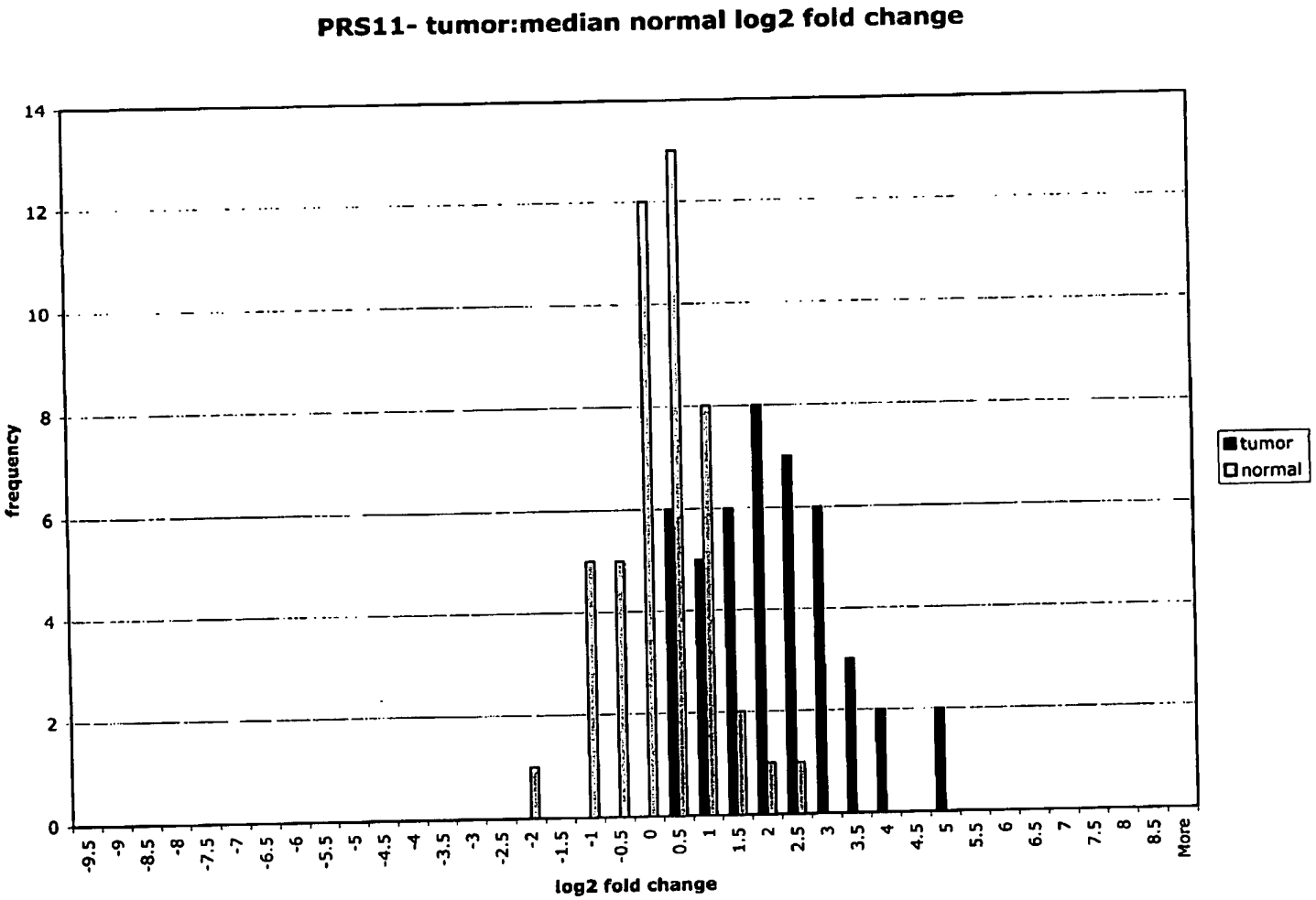
PRS11- tumor:median normal log2 fold change
frequency
log2 fold change
tumor
normal
0
2
4
6
8
10
12
14
-9.5
-9
-8.5
-8
-7.5
-7
-6.5
-6
-5.5
-5
-4.5
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
8
8.5
More

Figure 5(n)

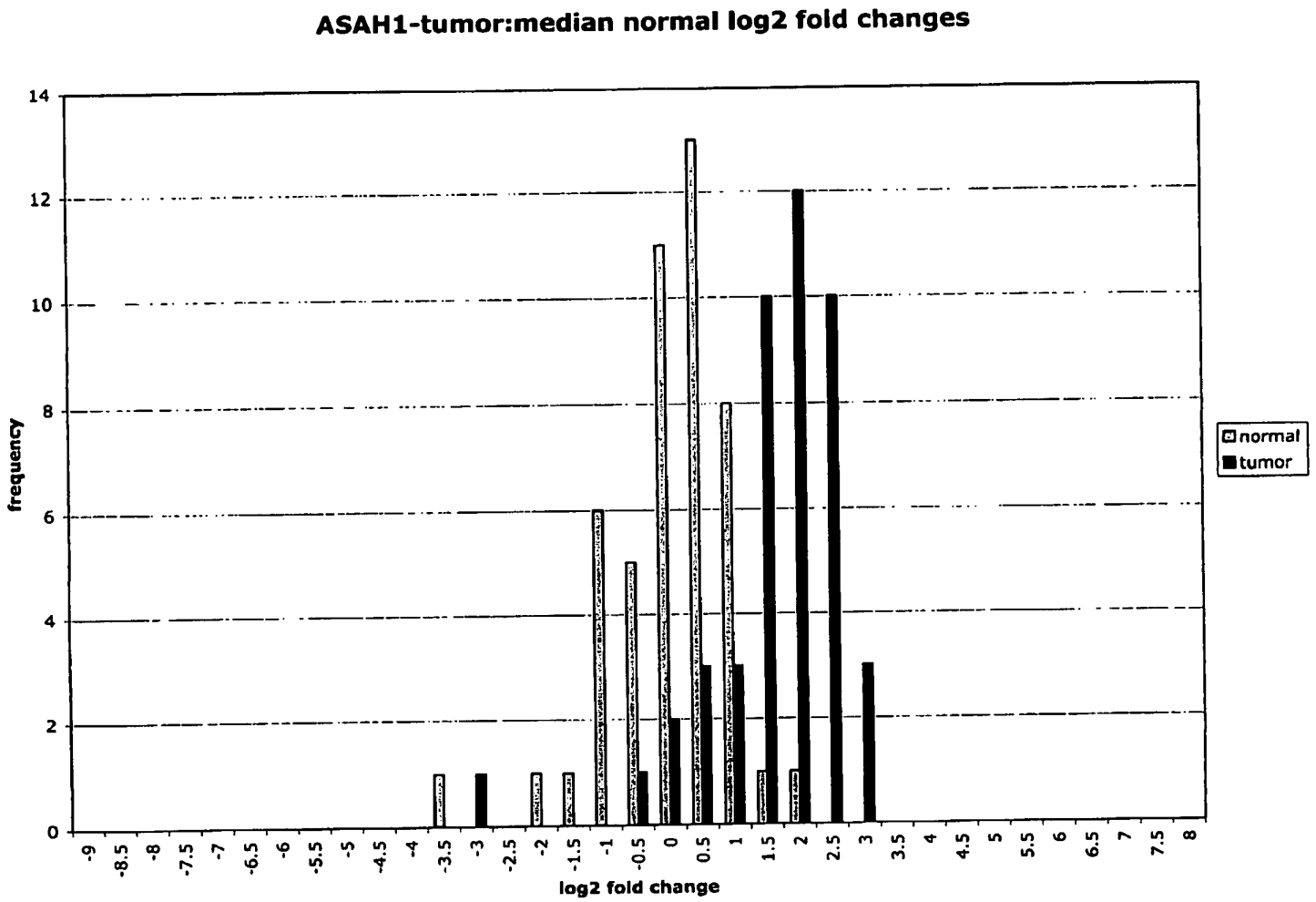
ASAH1-tumor:median normal log2 fold changes
frequency
log2 fold change
normal
tumor

Figure 5(o)

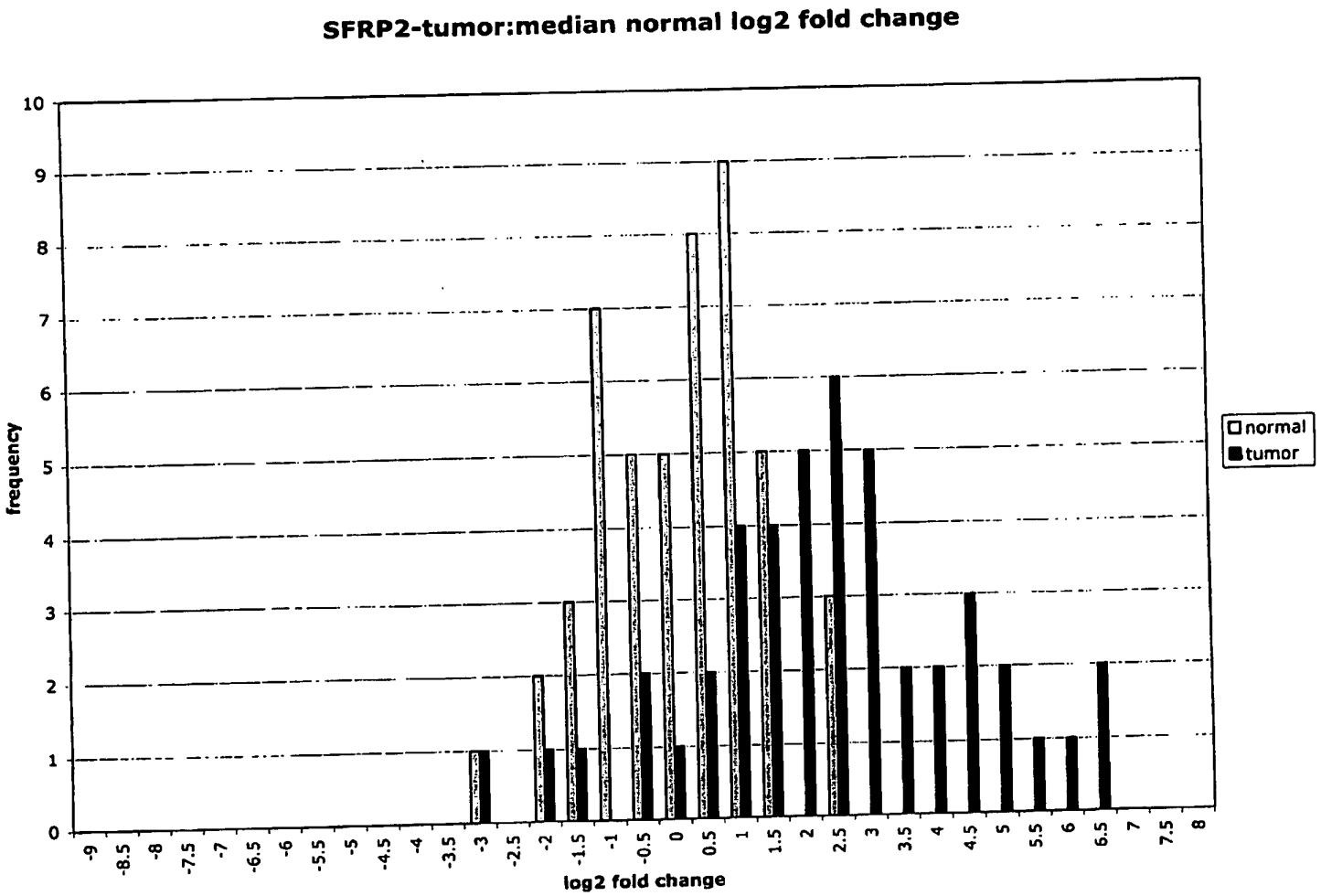
SFRP2-tumor:median normal log2 fold change
frequency
log2 fold change
normal
tumor

Figure 5(p)

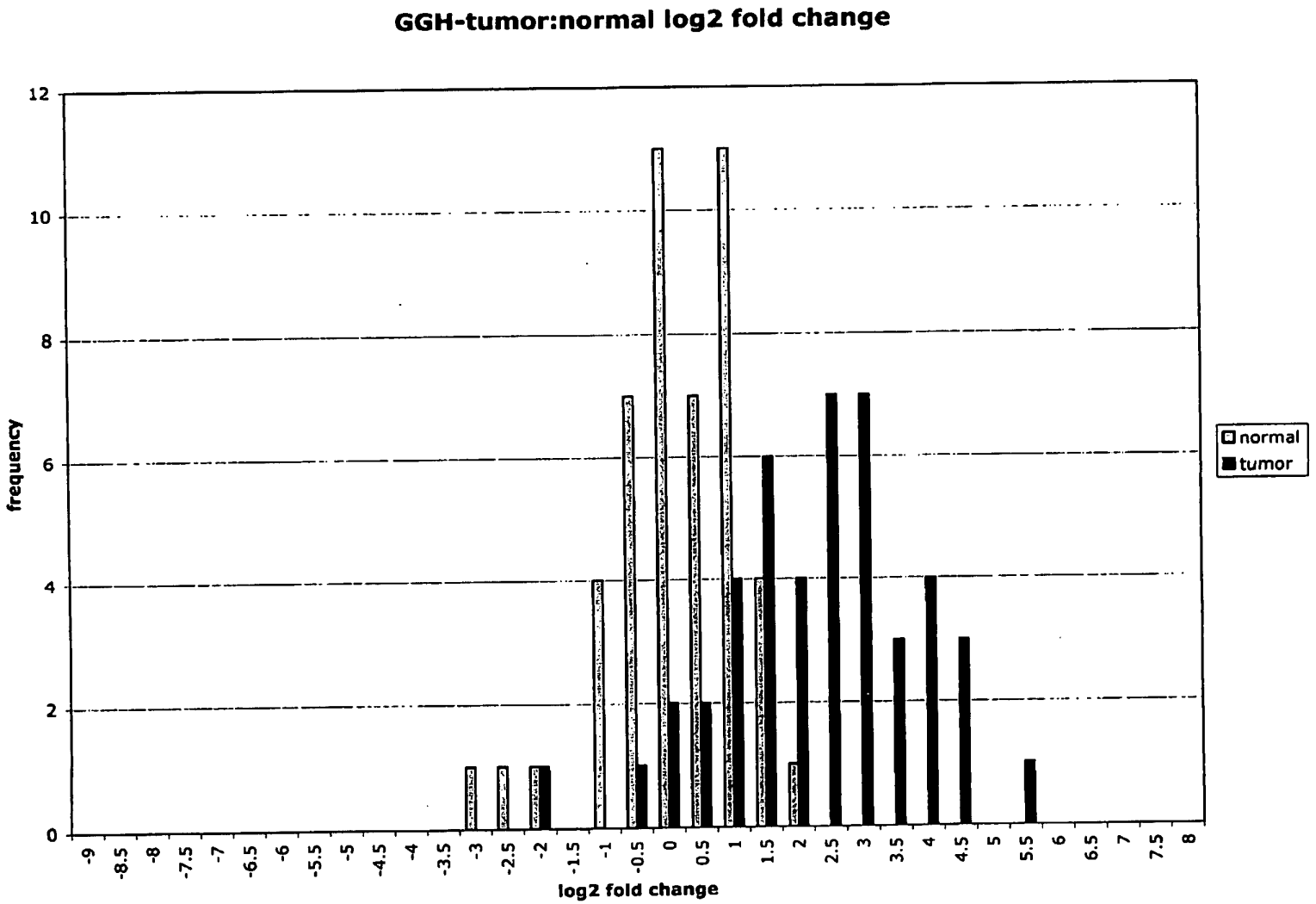
GGH-tumor:normal log2 fold change
frequency
log2 fold change
normal
tumor
0
2
4
6
8
10
12
-9
-8.5
-8
-7.5
-7
-6.5
-6
-5.5
-5
-4.5
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
8

Figure 5(q)

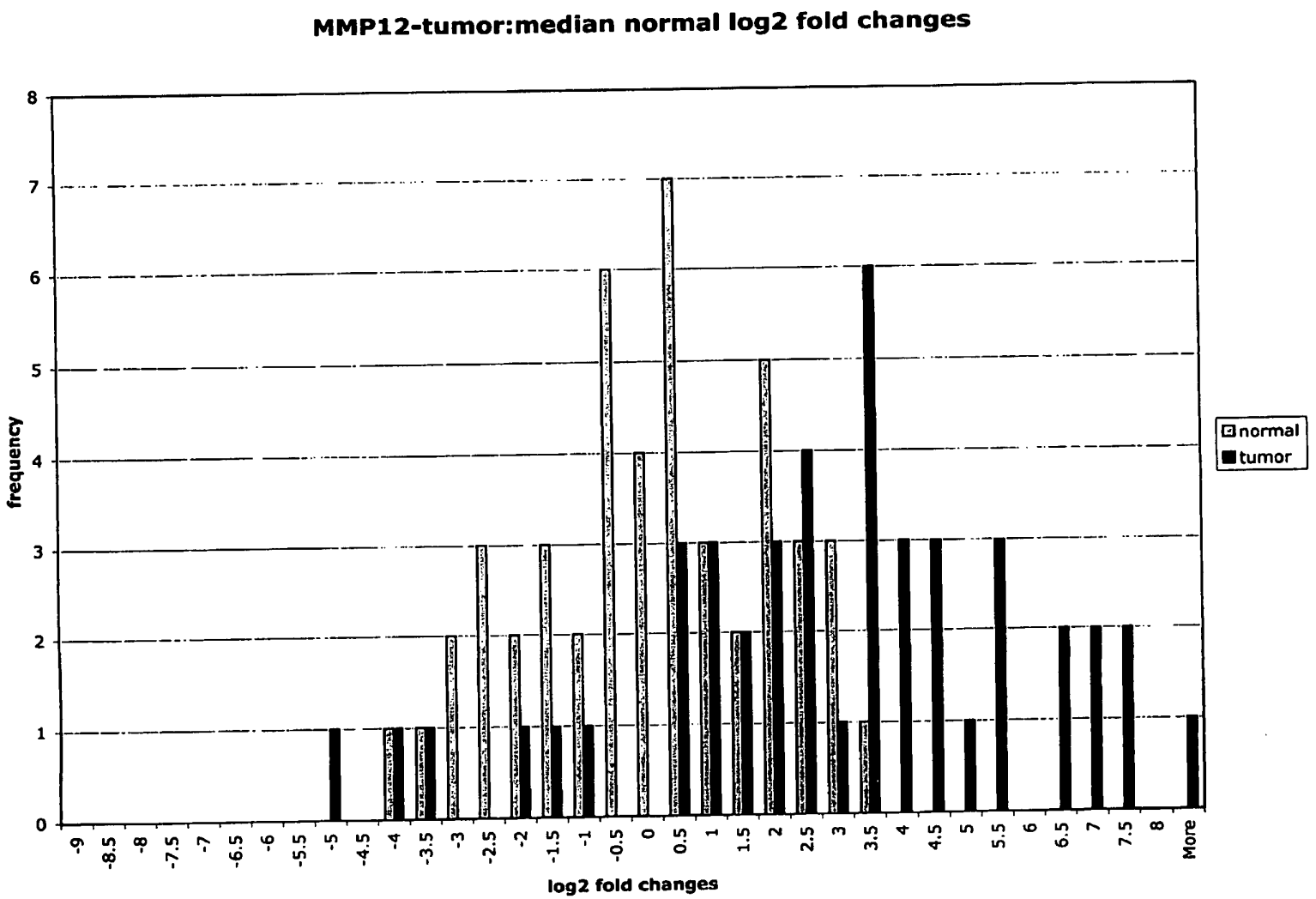
MMP12-tumor:median normal log2 fold changes
frequency
log2 fold changes
normal
tumor
0
1
2
3
4
5
6
7
8
-9
-8.5
-8
-7.5
-7
-6.5
-6
-5.5
-5
-4.5
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
8
More

Figure 5(r)

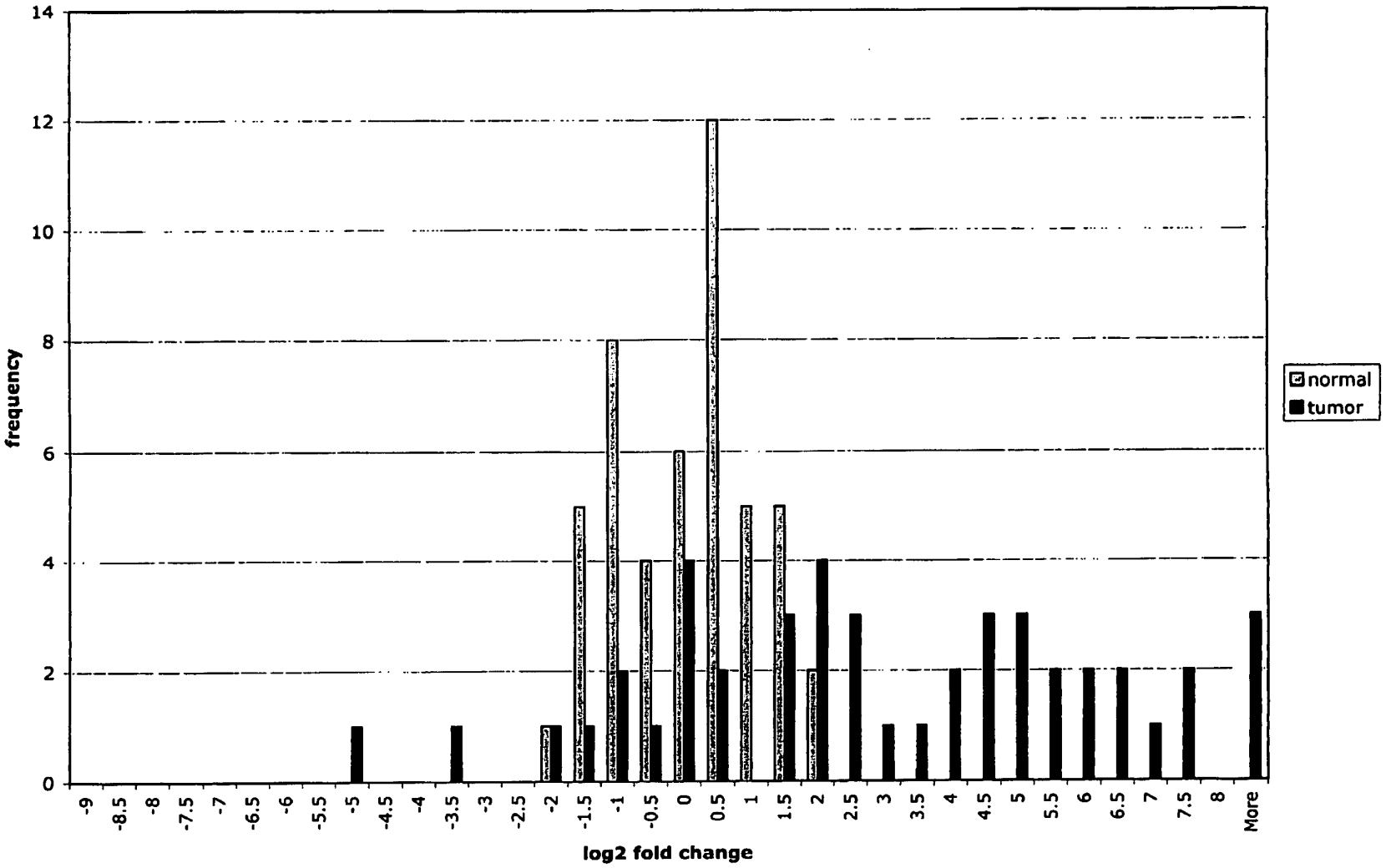
14
12
10
8
6
4
2
0
frequency
normal
tumor
-9
-8.5
-8
-7.5
-7
-6.5
-6
-5.5
-5
-4.5
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
8
More
log2 fold change

Figure 5(s)

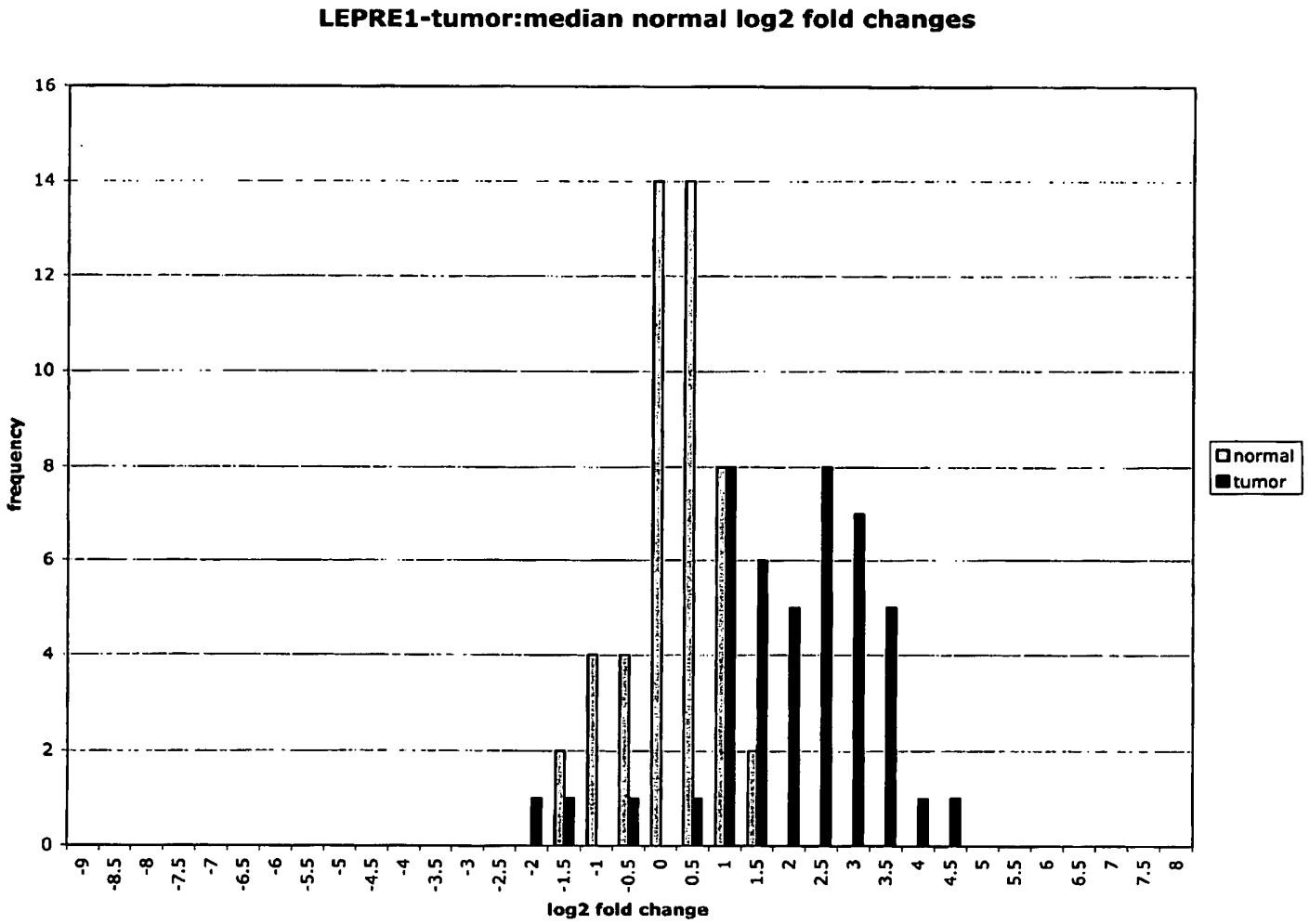
LEPRE1-tumor:median normal log2 fold changes
frequency
log2 fold change
normal
tumor

Figure 5(t)

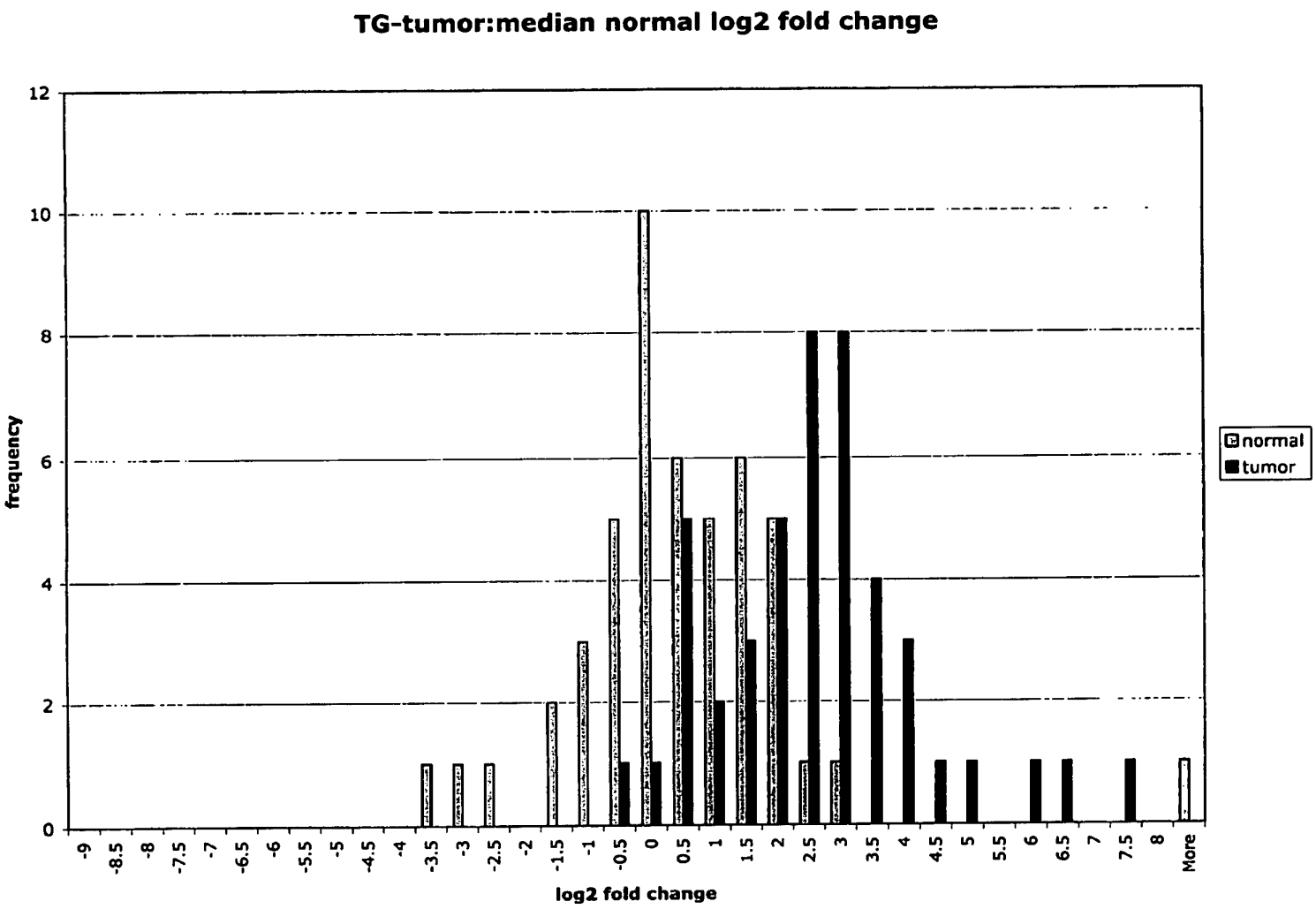
TG-tumor:median normal log2 fold change
frequency
log2 fold change
normal
tumor
0
2
4
6
8
10
12
-9
-8.5
-8
-7.5
-7
-6.5
-6
-5.5
-5
-4.5
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
8
More

Figure 5(u)

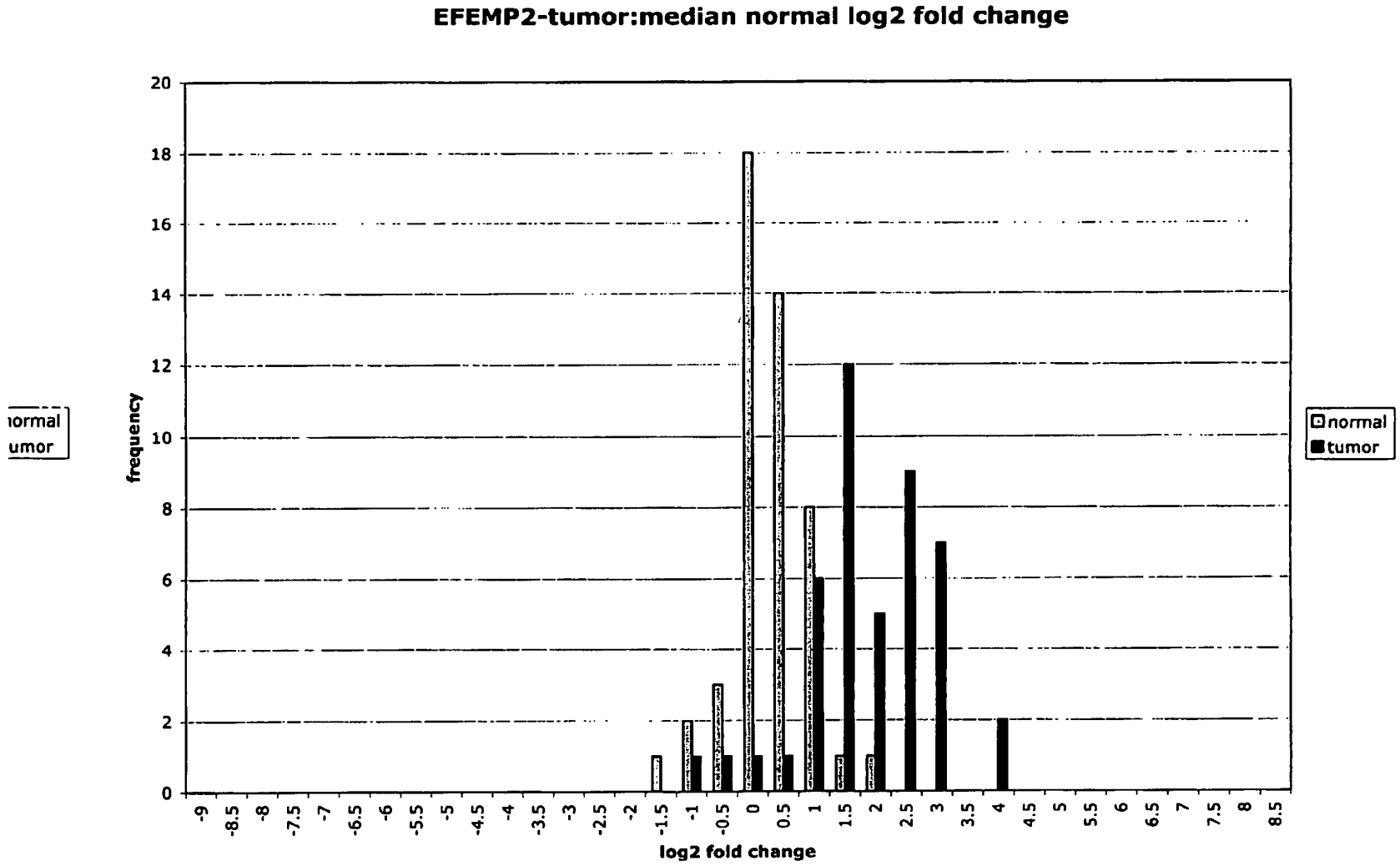
EFEMP2-tumor:median normal log2 fold change
frequency
log2 fold change
normal
tumor
0
2
4
6
8
10
12
14
16
18
20
-9
-8.5
-8
-7.5
-7
-6.5
-6
-5.5
-5
-4.5
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
8
8.5

Figure 5(v)

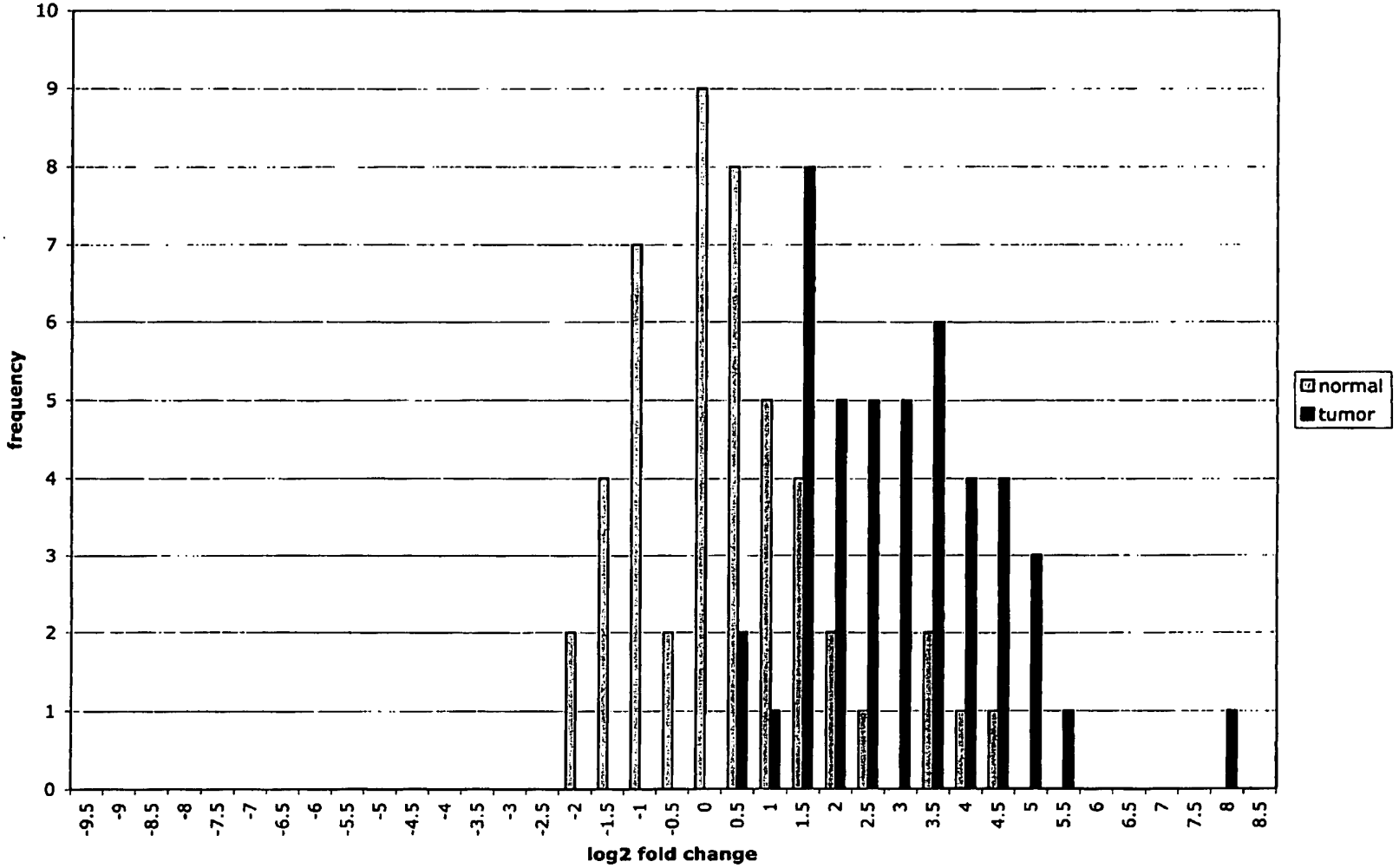

Figure 5(w)

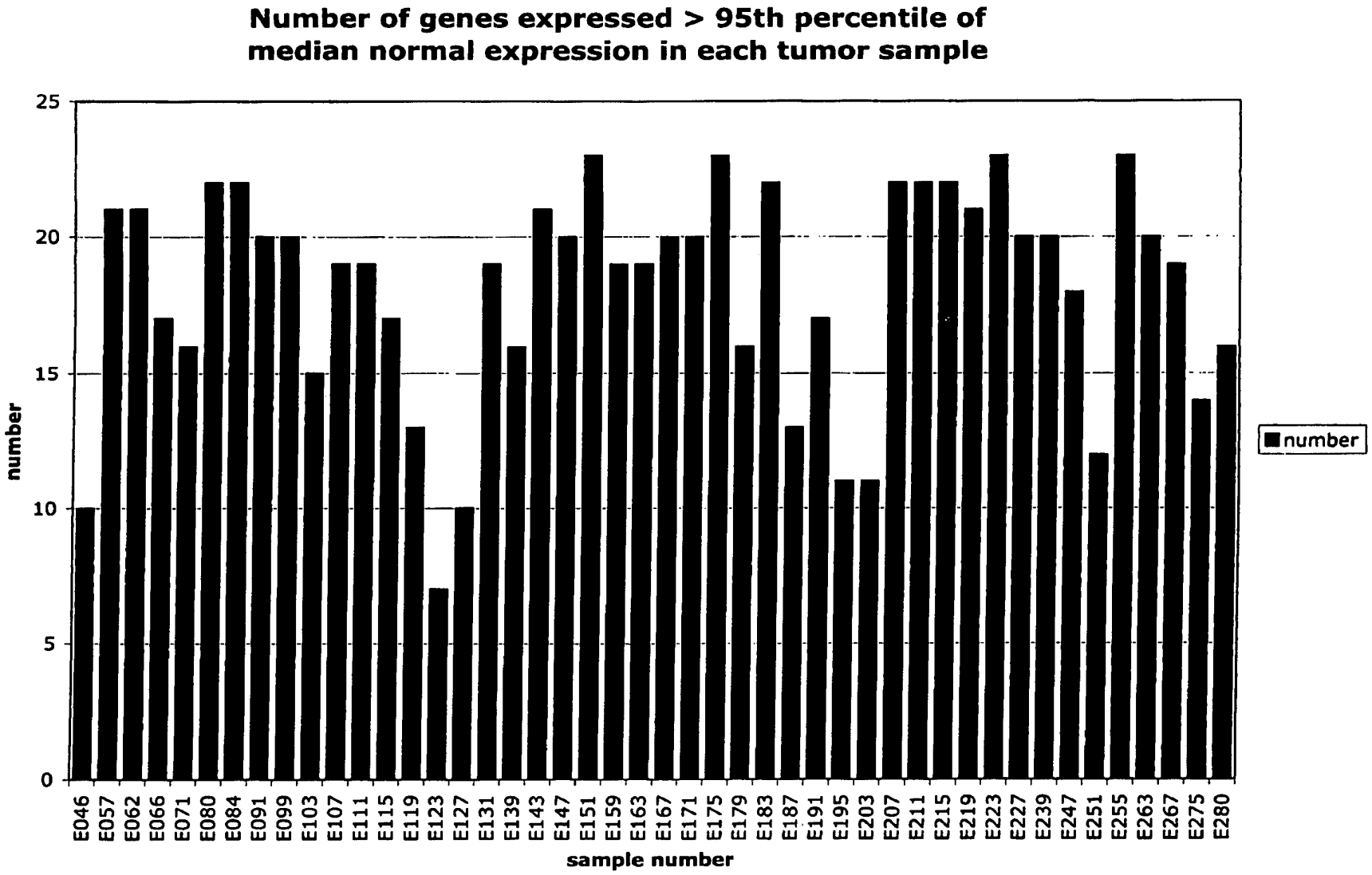
Number of genes expressed > 95th percentile of median normal expression in each tumor sample
number
sample number
number
0
5
10
15
20
25
E046
E057
E062
E066
E071
E080
E084
E091
E099
E103
E107
E111
E115
E119
E123
E127
E131
E139
E143
E147
E151
E159
E163
E167
E171
E175
E179
E183
E187
E191
E195
E203
E207
E211
E215
E219
E223
E227
E239
E247
E251
E255
E263
E267
E275
E280

Figure 6